When does partial commutative closure preserve regularity?

Antonio Cano Gómez* Giovanna Guaiana† Jean-Éric Pin‡§

To appear at ICALP 2008

The closure of a regular language under commutation or partial commutation has been extensively studied [1, 11, 12, 13], notably in connection with regular model checking [2, 3, 7] or in the study of Mazurkiewicz traces, one of the models of parallelism [14, 15, 16, 22]. We refer the reader to the survey [10, 9] or to the recent articles of Ochmański [17, 18, 19] for further references.

In this paper, we present new advances on two problems of this area. The first problem is well-known and has a very precise statement. The second problem is more elusive, since it relies on the somewhat imprecise notion of robust class. By a *robust class*, we mean a class of **regular languages** closed under some of the usual operations on languages, such as Boolean operations, product, star, shuffle, morphisms, inverses of morphisms, residuals, etc. For instance, regular languages form a very robust class, *commutative languages* (languages whose syntactic monoid is commutative) also form a robust class. Finally, *group languages* (languages whose syntactic monoid is a finite group) form a semi-robust class: they are closed under Boolean operation, residuals and inverses of morphisms, but not under product, shuffle, morphisms or star.

Here are the two problems:

Problem 1. *When is the closure of a regular language under [partial] commutation still regular?*

Problem 2. *Are there any robust classes of languages closed under [partial] commutation?*

The classes considered in this paper are all closed under polynomial operations. Recall that, given a class $\mathcal{L}$ of regular languages, the *polynomial languages* of $\mathcal{L}$ are the finite unions of languages of the form $L_0a_1L_1 \cdots a_kL_k$ where $a_1, \ldots, a_k$ are letters and $L_0, \ldots, L_k$ are languages of $\mathcal{L}$. Taking the polynomial closure usually increase robustness. For instance, the class $\mathrm{Pol}(\mathcal{G})$ of polynomials of group languages is closed under union, intersection, quotients, product, shuffle and inverses of morphisms.

Let I be a partial commutation and let D be its complement in $A \times A$. Our main results on Problems 1 and 2 can be summarized as follows:

*Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia, Camino de Vera s/n, P.O. Box: 22012, E-46020 - Valencia.

†LITIS EA 4108, Université de Rouen, BP12, 76801 Saint Etienne du Rouvray, France.

‡LIAFA, Université Paris-Diderot and CNRS, Case 7014, 75205 Paris Cedex 13, France.

§The authors acknowledge support from the AutoMathA programme of the European Science Foundation.

(1) The class $\mathrm{Pol}(\mathcal{G})$ is closed under commutation. If D is transitive, it is also closed under I-commutation.
(2) Under some simple conditions on the graph of I, the closure of a language of $\mathrm{Pol}(\mathcal{G})$ under I is regular.
(3) There is a very robust class of languages $\mathcal{W}$ which is closed under commutation. This class, which contains $\mathrm{Pol}(\mathcal{G})$, is closed under intersection, union, shuffle, concatenation, residual, length preserving morphisms and inverses of morphisms. Further, it is decidable and can be defined as the largest positive variety of languages not containing.
(4) If I is transitive, the closure of a language of $\mathcal{W}$ under I is regular.

Result (3) is probably the most important of the four results. It is, in a sense, optimal since $(ab)^*$ is the canonical example of a regular language whose commutative closure is not regular.

The proofs are nontrivial and combine several advanced techniques, including combinatorial Ramsey type arguments, algebraic properties of the syntactic monoid [5, 6], finiteness conditions on semigroups [8] and properties of insertion systems [4]. Some proofs are missing but are given in the Appendix.

Our paper is organised as follows. We first survey the known results in Section 1. Then we establish some combinatorial properties of group languages in Section 2. Our results on commutative closure are established in Section 3 and those on closure under partial commutation in Section 4.

1 Known results

Let A be an alphabet and let I be a symmetric and irreflexive relation on A (often called the *independence relation*). We denote by $\sim_I$ the congruence on A^* generated by the set $\{ab = ba \mid (a, b) \in I\}$. If L is a language on A^*, we also denote by $[L]_I$ the closure of L under $\sim_I$. When I is the relation $\{(a, b) \in A \times A \mid a \neq b\}$, we simplify the notation to $\sim$ and $[L]$, respectively. Thus $\sim$ is the commutation relation and $[L]$ is the *commutative closure* of L. The *dependence relation* associated with I is the relation $D = \{(a, b) \in A \times A \mid (a, b) \notin I\}$. The relations I and D define two graphs (A, I) and (A, D) with A as set of vertices. A class $\mathcal{C}$ of languages is *closed under I-commutation* if $L \in \mathcal{C}$ implies $[L]_I \in \mathcal{C}$.

1.1 The first problem

For the commutative closure, the problem is solved [11, 12, 13]: the commutative closure of the language $(ab)^*$ is not regular, but one can effectively decide whether the commutative closure of a regular language is regular or not.

For partial commutations, the result of Sakarovitch [22] concluded a series of previous partial results.

Theorem 1.1 *One can decide whether the closure $[L]_I$ of a regular language L is regular if and only if I is a transitive relation.*

1.2 The second problem

Only a few results are known for the second problem. They concern the following classes of languages:

(1) the class $\mathrm{Pol}(\mathcal{I})$ of finite unions of languages of the form $A^*a_1A^* \cdots a_kA^*$, with $a_1, \ldots, a_k \in A$,

(2) the class $\mathcal{J}$ of piecewise testable languages (the Boolean closure of $\mathrm{Pol}(\mathcal{I})$),

(3) the class $\mathrm{Pol}(\mathcal{J})$, which consists of finite unions of languages of the form $A_0^*a_1A_1^* \cdots a_kA_k^*$ with $A_i \subseteq A$ and $a_1, \ldots, a_k \in A$. These languages are also called *APC* (*Alphabetic Pattern Constraints*) in [2],

(4) the class $\mathrm{Pol}(\mathcal{C}om)$ of polynomials of commutative languages.

The following theorem summarises the results of Guaiana, Restivo and Salemi [14, 15], Bouajjani, Muscholl and Touili [2, 3] and Cécé, Héam and Mainier [7].

Theorem 1.2 *Let I be any independence relation. Then*

(1) *the class $Pol(\mathcal{I})$ is closed under commutation,*

(2) *the class $\mathcal{J}$ is closed under commutation,*

(3) *the class $Pol(\mathcal{J})$ is closed under I-commutation,*

(4) *the class $Pol(\mathcal{C}om)$ is closed under I-commutation.*

1.3 Star-free languages

Two nice results on star-free languages were proved by Muscholl and Petersen [16]. The first one is the counterpart of Theorem 1.1 for star-free languages.

Theorem 1.3 *One can decide whether the closure $[L]_I$ of a star-free language L is star-free if and only if I is a transitive relation.*

The second result is related to our second problem.

Theorem 1.4 *Let L be a star-free language. If D is transitive, then $[L]_I$ is either star-free or non regular. If D is not transitive, then there exist star-free languages such that $[L]_I$ is regular but not star-free.*

2 Properties of group languages

Recall that a *group language* is a regular language whose syntactic monoid is a group, or, equivalently, is recognized by a finite deterministic automaton in which each letter defines a permutation of the set of states. We gather in this section the properties of these languages needed in this paper.

Let π be a morphism from A^* onto a finite group G and let $R = \pi^{-1}(1)$. The next lemma is a well-known consequence of Ramsey's theorem [20].

Lemma 2.1 *For any $n > 0$, there exists $N > 0$ such that, for any u_0, u_1, $\ldots$, $u_N \in A^*$ there exists a sequence $0 \leqslant i_0 < i_1 < \ldots < i_n \leqslant N$ such that $\pi(u_{i_0}u_{i_0+1} \cdots u_{i_1-1}) = \pi(u_{i_1}u_{i_1+1} \cdots u_{i_2-1}) = \ldots = \pi(u_{i_{n-1}} \cdots u_{i_n-1}) = 1$.*

An *insertion system* is a special type of rewriting system whose rules are of the form $1 \to r$ for all r in a given language R. We write $u \to_R v$ if $u = u'u''$ and $v = u'ru''$ for some $r \in R$. We denote by $\xrightarrow{*}_R$ the reflexive transitive closure of the relation $\to_R$. Given a language L of A^*, its closure under $\xrightarrow{*}_R$ is the language

$$[L]_{\xrightarrow{*}_R} = \{v \in A^* \mid \text{ there exists } u \in L \text{ such that } u \xrightarrow{*}_R v.\}$$

We are especially interested in the case $R = \pi^{-1}(1)$, where π is a morphism from A^* onto a finite group G. Let F be the set of words of R of length $\leqslant |G|$. It is not difficult to see that the relations $\xrightarrow{*}_F$ and $\xrightarrow{*}_R$ coincide. The next result follows from the results of Bucher, Ehrenfeucht and Haussler [4].

Proposition 2.2

(1) *The relation $\leqslant_\pi$ is a well preorder on A^*.*

(2) *For any language L, the language $[L]_{\xrightarrow{*}_R}$ is regular.*

We prove a slightly more precise result.

Proposition 2.3 *For any language L, the language $[L]_{\xrightarrow{*}_R}$ is a polynomial of group languages.*

Proof. Let $R = \pi^{-1}(1)$. By construction, R is a group language. If $u = a_1 \cdots a_n$, the language $[u]_{\xrightarrow{*}_R}$ is equal to $Ra_1R \cdots Ra_nR$, a polynomial of group languages. Now since $\leqslant_\pi$ is a well preorder, every language of the form $[L]_{\xrightarrow{*}_R}$ is equal to a language of the form $[F]_{\xrightarrow{*}_R}$ with F finite and thus a finite union of languages of the form $Ra_1R \cdots Ra_nR$. It is therefore a polynomial of group languages. □

3 Commutative closure

This section contains three new results. The first one concerns group languages, the second one polynomials of group languages and the third one a robust class introduced in [5, 6] and denoted by $\mathcal{W}$.

Recall that if L is a language of A^*, the *syntactic preorder of* L is the relation $\leqslant_L$ defined on A^* by $u \leqslant_L v$ if and only if, for every $x, y \in A^*$, $xvy \in L$ implies $xuy \in L$. The syntactic congruence $\sim_L$ is defined by $u \sim_L v$ if and only if $u \leqslant_L v$ and $v \leqslant_L u$.

3.1 Group languages

Theorem 3.1 *The commutative closure of a group language is regular.*

Proof. Let $L \subseteq A^*$ be a group language and let $\pi : A^* \to G$ be its syntactic morphism. Let $n = |G|$ and let N be the integer given by Lemma 2.1. We claim that for any letter $a \in A$, $a^N \sim_{[L]} a^{N+n}$. Let $g = \pi(a)$.

Suppose that $xa^N y \in [L]$. Then there exists a word w of L commutatively equivalent to $xa^N y$. It follows that wa^n is commutatively equivalent to $xa^{N+n}y$. Further, since G is a finite group, one has $g^n = 1$ by Lagrange's theorem, whence $\pi(wa^n) = \pi(w)\pi(a^n) = \pi(w)$. Thus the words w and wa^n have the same syntactic image by π and hence $wa^n \in L$. Therefore $xa^{N+n}y \in [L]$.

Conversely, assume that $xa^{N+n}y \in [L]$. Then $xa^{N+n}y$ is commutatively equivalent to some word of L, say $w = u_0au_1a \cdots u_{N-1}au_Nau_{N+1}$. By applying Lemma 2.1 to the sequence of words u_0a, u_1a, $\ldots$, u_Na, we obtain a sequence $0 \leqslant i_0 < i_1 < \ldots < i_n \leqslant N$ such that

$$\pi(u_{i_0}a \cdots au_{i_1-1}a) = \pi(u_{i_1}a \cdots au_{i_2-1}a) = \ldots = \pi(u_{i_{n-1}}a \cdots au_{i_n-1}a) = 1 \quad (1)$$

This implies in particular

$$\pi(u_{i_0} a \cdots a u_{i_1 - 1}) = \pi(u_{i_1} a \cdots a u_{i_2 - 1}) = \ldots = \pi(u_{i_{n-1}} a \cdots a u_{i_n - 1}) = g^{-1} \quad (2)$$

Let r and s be the words defined by

$$w = r(u_{i_0} a \cdots a u_{i_1 - 1} a)(u_{i_1} a \cdots a u_{i_2 - 1} a)(u_{i_{n-1}} a \cdots a u_{i_n - 1} a)s$$

Since w is commutatively equivalent to $xa^{N+n}y$, the word

$$w' = r(u_{i_0} a \cdots a u_{i_1 - 1})(u_{i_1} a \cdots a u_{i_2 - 1}) \cdots (u_{i_{n-1}} a \cdots a u_{i_n - 1})s$$

is commutatively equivalent to $xa^N y$. Furthermore, Formulas (1) and (2) show that $\pi(w) = \pi(r)\pi(s)$ and $\pi(w') = \pi(r)(g^{-1})^n \pi(s)$. Since $(g^{-1})^n = 1$ by Lagrange's theorem, $\pi(w) = \pi(w')$ and thus $w' \in L$. It follows that $xa^N y \in [L]$, which proves the claim.

Now, the syntactic monoid of $[L]$ is a commutative monoid in which each generator has a finite index. Since the alphabet is finite, this monoid is finite and thus $[L]$ is regular. $\square$

Theorem 3.1 indicates that the commutative closure of a group language is a commutative regular language. One may wonder whether, in turn, any commutative regular language is the commutative closure of a group language. The answer is no, but requires an improved version of Theorem 3.1.

Theorem 3.2 *The commutative closure of a group language is a polynomial of group languages.*

Proof. Let L be a group language and let $\pi : A^* \to G$ be its syntactic morphism. We claim that $[L]$ is a filter for $\leqslant_\pi$, which will give the result by Proposition 2.3. Let us show that if $a_1 \cdots a_n \in [L]$ and $v_0, v_1, \ldots, v_n \in \pi^{-1}(1)$, then $v_0 a_1 v_1 \cdots a_n v_n \in [L]$. Since $a_1 \cdots a_n \in [L]$, there exists a word $w \in L$ which is commutatively equivalent to $a_1 \cdots a_n$. Thus the word $wv_0 v_1 \cdots v_n$ is commutatively equivalent to $v_0 a_1 v_1 \cdots a_n v_n$. Now $\pi(wv_0 v_1 \cdots v_n) = \pi(w)\pi(v_0) \cdots \pi(v_n) = \pi(w)$. Therefore $wv_0 v_1 \cdots v_n \in L$, proving the claim. $\square$

Note that the commutative closure of a group language is not necessarily a group language. Indeed, consider the set of all words of $\{a, b\}^*$ having an even number of (scattered) subwords equal to ab. Its commutative closure, $A^*aA^*bA^* \cup A^*bA^*aA^*$ is not a group language. However, Theorem 3.2 can be extended to polynomials of group languages.

Theorem 3.3 *The commutative closure of a polynomial of group languages is also a polynomial of group languages.*

Proof. It is shown in [21] that for any polynomial of group languages L, there exists a morphism $\pi : A^* \to G$ from A^* onto a finite group G such that L is a finite union of languages of the form $Ra_1R \cdots Ra_nR$, with $R = \pi^{-1}(1)$. Thus it suffices to show that if $K = Ra_1R \cdots Ra_nR$ for some letters $a_1, \ldots, a_n$, then $[K]$ is a polynomial of group languages.

We claim that $[K]$ is a filter for $\leqslant_\pi$, which will give the result by Proposition 2.3. Let us show that if $b_1 \cdots b_m \in [K]$ and $v_0, v_1, \ldots, v_n \in R$, then $v_0 b_1 v_1 \cdots b_m v_n \in [K]$. Let w be a word of K commutatively equivalent to

$b_1 \cdots b_m$. As an element of K, w can be written as $r_0 a_1 r_1 \cdots a_n r_n$ for some words $r_0, \ldots, r_n \in R$. Since the words $v_0, \ldots, v_m$ are in R, the word $w v_0 v_1 \cdots v_n$ also belongs to K and is commutatively equivalent to $v_0 b_1 v_1 \cdots b_m v_m$. This proves the claim and concludes the proof. □

3.2 Languages of $\mathcal{W}$

We now define the class of regular languages $\mathcal{W}$ first introduced and studied in [5, 6]. Recall that a *positive variety* of languages is a class of regular languages closed under union, intersection, residuals and inverses of morphisms.

The class $\mathcal{W}$ is the unique maximal positive variety of languages which does not contain the language $(ab)^*$, for all letters $a \neq b$. It is also the unique maximal positive variety satisfying the two following conditions: it is *proper*, that is, strictly included in the variety of regular languages, and it is closed under the shuffle operation. It is also the largest proper positive variety closed under length preserving morphisms. Being closed under intersection, union, shuffle, concatenation, length preserving morphisms and inverses of morphisms, $\mathcal{W}$ is a quite robust class, which strictly contains the classes APC, Pol($\mathcal{C}om$) and Pol($\mathcal{G}$) introduced in Section 1.2.

The class $\mathcal{W}$ has an algebraic characterization [5, 6]. Let a and b be two elements of a monoid. Recall that b is an *inverse of* a if $aba = a$ and $bab = b$. Now, a regular language belongs to $\mathcal{W}$ if and only if its syntactic ordered monoid belongs to the variety of finite ordered monoids $\mathbf{W}$ defined as follows: an ordered monoid $(M, \leqslant)$ belongs to $\mathbf{W}$ if and only if, for any pair (a, b) of mutually inverse elements of M, and any element z of the minimal ideal of the submonoid generated by a and b, $(abzab)^\omega \leqslant ab$ (see [6, p.435–436] for a precise definition of the semigroup notions used in this characterization). This description might appear quite involved, but has an important consequence: the variety $\mathcal{W}$ is decidable. That is, given a regular language L, one can decide whether or not L belongs to $\mathcal{W}$. We also mention for the specialists that $\mathbf{W}$ contains the variety of finite monoids $\mathbf{DS}$.

The main result of this section states that $\mathcal{W}$ is closed under commutative closure. In fact, we prove a stronger result, which relies on the notion of a period that we now introduce.

Let M be a finite monoid. The *exponent* of M is the least integer ω such that for all $x \in M$, x^ω is idempotent. Its *period* is the least integer p such that for all $x \in M$, $x^{\omega+p} = x^\omega$. By extension, the *period* (respectively *exponent*) of a regular language is the period (respectively exponent) of its syntactic monoid.

Proposition 3.4 *Let L be a commutative language of A^* and let d be a positive integer. If, for each letter c of A, there exists $N > 0$ such that $c^{N+d} \leqslant_L c^N$, then L is regular and its period divides d.*

Proof. It follows from [8, Theorem 6.6.2, page 215] that, under these conditions, L is a regular language. Let ω be the exponent of L. The relation $c^{N+d} \leqslant_L c^N$ gives $c^{N(\omega-1)} c^{N+d} \leqslant_L c^{N(\omega-1)} c^N$, whence $c^{N\omega+d} \leqslant_L c^{N\omega}$ and since $c^\omega \sim_L c^{2\omega} \sim_L c^{N\omega}$, one gets finally $c^{\omega+d} \leqslant_L c^\omega$. It follows that

$$c^\omega \sim_L c^{\omega+\omega d} \leqslant_L \ldots \leqslant_L c^{\omega+2d} \leqslant_L c^{\omega+d} \leqslant_L c^\omega$$

and hence $c^\omega \sim_L c^{\omega+d}$. Since L is commutative, its syntactic monoid is commutative and therefore $u^\omega \sim_L u^{\omega+d}$ for all $u \in A^*$. It follows that the period of L divides d. $\square$

We can now state:

Theorem 3.5 *Let L be a language of $\mathcal{W}(A^*)$. Then $[L]$ is regular and commutative (and hence in $\mathcal{W}$) and its period divides that of L.*

Proof. Let L be a language of $\mathcal{W}(A^*)$ and let $[L]$ be its commutative closure. Then there exist an ordered monoid $(M, \leqslant) \in \mathbf{W}$, a surjective monoid morphism $\varphi : A^* \to M$ and an order ideal P of $(M, \leqslant)$ such that $\varphi^{-1}(P) = L$. Let ω be the exponent of M and let p be its period. Let also d be any number such that, for all $t \in M$, t^d is idempotent. In particular, d can be either ω or $\omega + p$. We claim that, for every such d, there exists an integer N such that, for every letter $c \in A$, $c^{N+d} \leqslant_{[L]} c^N$. If the claim holds, then Proposition 3.4 shows that $[L]$ is regular and that its period divides d. Taking $d = \omega$ and $d = \omega + p$ then proves that this period also divides p.

The rest of the proof consists in proving the claim. We need two combinatorial results. The first one is a slight variation of Lemma 2.1.

Lemma 3.6 *Let c be a letter of A. For any $n \geqslant 0$, there exists an integer N such that, for every word u of A^* containing at least $N+1$ occurrences of c, there exist an idempotent e of M and a factorization $u = v_0v_1cv_2c\cdots v_ncv_{n+1}$ such that, for $1 \leqslant i \leqslant n$, $\varphi(v_ic) = e$.*

The second one requires an auxiliary definition. A word u of $\{a, b\}^*$ is said to be *balanced* if $|u|_a = |u|_b$.

Proposition 3.7 *Let $B = \{a, b\}$. There exists a balanced word $z \in B^*$ such that, for any morphism $\gamma : B^* \to M$, $\gamma(z)$ belongs to the minimal ideal of the monoid $\gamma(B^*)$.*

Proof. Let $n = |M|$ and let z be a balanced word of B^* containing all words of length $\leqslant n$ as a factor. Let $\gamma : B^* \to M$ be a morphism and let m be an element of the minimal ideal J of $\gamma(B^*)$. Then one can show there is a word u of length $\leqslant n$ such that $\gamma(u) = m$. Since $|u| \leqslant n$, u is a factor of z and $\gamma(z)$ belongs to $M\gamma(u)M$. Now since $m \in J$, $M\gamma(u)M = MmM = J$ and hence $\gamma(z) \in J$. $\square$

Let us continue the proof of Theorem 3.5. Let $n = |M|$ and let z be the balanced word given by Proposition 3.7. Let $r = |z|_a = |z|_b$, $n_3 = d(1 + r)$, $n_2 = nn_3$ and $n_1 = 3n_2$. Finally let $N = N(n_1)$ be the constant given by Lemma 3.6.

Let $x, y \in A^*$. If $xc^Ny \in [L]$, there exists a word u of L commutatively equivalent to xc^Ny and hence containing at least N occurrences of c. By Lemma 3.6, there exist an idempotent e of M and a factorization $u = v_0v_1c \ \cdots \ v_{n_1}cv_{n_1+1}$ such that, for $1 \leqslant i \leqslant n_1$, $\varphi(v_ic) = e$.

Now, since $n_1 = 3n_2$, one can also write u as $u = v_0(f_1g_1) \ \cdots \ (f_{n_2}g_{n_2})v_{n_1+1}$ where, for $1 \leqslant i \leqslant n_2$, $f_i = v_{3i-2}cv_{3i-1}$ and $g_i = cv_{3i}c$.

Lemma 3.8 *For $1 \leqslant i \leqslant n_2$, the elements $\varphi(f_i)$ and $\varphi(g_i)$ are mutually inverse.*

Proof. We omit this proof, but it is a straightforward verification. $\square$

Setting $\bar{s} = \varphi(c)e$, one gets $\varphi(g_i) = \bar{s}$ for $1 \leqslant i \leqslant n_2$. Further, by the choice of n_2 and by the pigeonhole principle, one can find n_3 indices $i_1 < \ldots < i_{n_3}$ and an element $s \in M$ such that $\varphi(f_{i_1}) = \ldots = \varphi(f_{i_{n_3}}) = s$. Setting

$$\begin{aligned}
w_0 &= v_0 f_1 g_1 \ \cdots f_{i_1-1} g_{i_1-1} & x_1 &= f_{i_1} & y_1 &= g_{i_1} \\
w_1 &= f_{i_1+1} g_{i_1+1} \ \cdots \ f_{i_2-1} g_{i_2-1} & x_2 &= f_{i_2} & y_2 &= g_{i_2} \\
&\vdots & &\vdots & & \\
w_{n_3-1} &= f_{i_{n_3-1}+1} g_{i_{n_3-1}+1} \ \cdots \ f_{i_{n_3}-1} g_{i_{n_3}-1} & x_{n_3} &= f_{i_{n_3}} & y_{n_3} &= g_{i_{n_3}} \\
w_{n_3} &= f_{i_{n_3}+1} g_{i_{n_3}+1} \ \cdots \ f_{n_2} g_{n_2} v_{n_1+1} & & & &
\end{aligned}$$

we obtain a factorization

$$u = w_0 x_1 y_1 w_1 x_2 y_2 w_2 \ \cdots \ w_{n_3-1} x_{n_3} y_{n_3} w_{n_3} \tag{3}$$

such that $\varphi(w_1) = \ldots = \varphi(w_{n_3-1}) = e$, $\varphi(x_1) = \ldots = \varphi(x_{n_3}) = s$ and $\varphi(y_1) = \ldots = \varphi(y_{n_3}) = \bar{s}$.

Recall that $n_3 = d(1+r)$ where $r = |z|_a = |z|_b$. We now define words z_1, $\ldots$, z_d as follows: the word z_j is obtained by replacing in z the first occurrence of a by $x_{d+(j-1)r+1}$, the second occurrence of a by $x_{d+(j-1)r+2}$, $\ldots$, the r's occurrence of a by x_{d+jr} and, similarly, the first occurrence of b by $y_{d+(j-1)r+1}$, the second occurrence of b by $y_{d+(j-1)r+2}$, $\ldots$, the r's occurrence of b by y_{d+jr}. Finally, set

$$\begin{aligned}
u' = w_0(v_{3i_1-2} c c v_{3i_1-1} c z_1 v_{3i_1} c)(v_{3i_2-2} c c v_{3i_2-1} c z_2 v_{3i_2} c) \ \cdots \\
(v_{3i_d-2} c c v_{3i_d-1} c z_d v_{3i_d} c) w_1 \ \cdots \ w_{n_3}
\end{aligned} \tag{4}$$

We are now ready for the three final steps.

Lemma 3.9 *The word u' is commutatively equivalent to $xc^{N+d}y$.*

Proof. It is clear that u' is commutatively equivalent to

$$c^d w_0 (v_{3i_1-2} c v_{3i_1-1} c v_{3i_1} c) \ \cdots (v_{3i_d-2} c v_{3i_d-1} c v_{3i_d} c)(z_1 \ \cdots \ z_d)(w_1 \ \cdots \ w_{n_3})$$

Now, $v_{3i_1-2} c v_{3i_1-1} c v_{3i_1} c = f_{i_1} g_{i_1} = x_1 y_1$, $\ldots$, $v_{3i_d-2} c v_{3i_d-1} c z_d v_{3i_d} c = f_{i_d} g_{i_d} = x_d y_d$. Further, by construction, $z_1 \ \cdots \ z_d \sim x_{d+1} y_{d+1} \ \cdots \ x_{n_3} y_{n_3}$. Therefore

$$u' \sim c^d w_0 x_1 y_1 w_1 x_2 y_2 w_2 \ \cdots \ w_{n_3-1} x_{n_3} y_{n_3} w_{n_3}$$

and finally $u' \sim uc^d \sim xc^{N+d}y$. $\square$

Let T be the submonoid of M generated by s and $\bar{s}$ and let $\gamma : \{a,b\}^* \to T$ be the morphism defined by $\gamma(a) = s$ and $\gamma(b) = \bar{s}$. By Proposition 3.7, $\gamma(z)$ belongs to the minimal ideal of T and since $e = s\bar{s}$, the definition of $\mathbf{W}$ shows that in M, $(e\gamma(z)e)^d \leqslant e$.

Lemma 3.10 *One has $\varphi(z_1) = \ldots = \varphi(z_d) = \gamma(z)$.*

Proof. Each of the words z_j is obtained by replacing in z the occurrences of a by some x_k and each occurrence of b by some y_k. Since all the x_k (resp. y_k) have the same image by φ, namely s (resp. $\bar{s}$), $\varphi(z_j)$ is equal to $\gamma(z)$. $\square$

Lemma 3.11 *The word u' belongs to L.*

Proof. It follows from (3) that $\varphi(u) = \varphi(w_0)e\varphi(w_{n_3})$, and hence, since $P = \varphi(L)$, $\varphi(w_0)e\varphi(w_{n_3}) \in P$. Now, observe that

$$\begin{aligned}\varphi(v_{3i_1-2}ccv_{3i_1-1}cz_1v_{3i_1}c) &= \varphi(v_{3i_1-2}c)\varphi(c)\varphi(v_{3i_1-1}c)\varphi(z_1)\varphi(v_{3i_1}c)\\ &= e\varphi(c)e\varphi(z_1)e = e\bar{s}\gamma(z)e \quad \text{by Lemma 3.10}\end{aligned}$$

By a similar argument, one has

$$\varphi(v_{3i_1-2}ccv_{3i_1-1}cz_1v_{3i_1}c) = \ldots = \varphi(v_{3i_d-2}ccv_{3i_d-1}cz_dv_{3i_d}c) = e\bar{s}\gamma(z)e$$

Finally, since $\varphi(w_1) = \ldots = \varphi(w_{n_3-1}) = e$, it follows from (4) that

$$\varphi(u') = \varphi(w_0)(e\bar{s}\gamma(z)e)^d\varphi(w_{n_3})$$

Furthermore, since $\bar{s} \in T$, $\bar{s}\gamma(z)$ belongs to the minimal ideal of T and since M is in $\mathbf{W}$, one has $(e\bar{s}\gamma(z)e)^d \leqslant e$. Since $\varphi(L)$ is an order ideal, the element $\varphi(w_0)(e\bar{s}\gamma(z)e)^d\varphi(w_{n_3})$ is also in $\varphi(L)$ and hence $u' \in L$. $\square$

Putting Lemmas 3.9 and 3.11 together, we conclude that $xc^{N+d}y \in [L]$, which proves the claim and the theorem. $\square$

4 Closure under partial commutation

Some of the results of Section 3 can be extended to partial commutations, under some restrictions on the set I.

4.1 The case where D is transitive

The condition that D is transitive is equivalent to requiring that $A^*/{\sim_I}$ is isomorphic to a direct product of free monoids $A_1^* \times \cdots \times A_k^*$. Denote by π_j the projection from A^* onto A_j^* and let π_I be the morphism from A^* onto $A_1^* \times \cdots \times A_k^*$ defined by $\pi_I(u) = (\pi_1(u), \ldots, \pi_k(u))$. This morphism is intimately connected to our problem, since $u \sim_I v$ if and only if $\pi_I(u) = \pi_I(v)$. Let us denote by Ш the shuffle product. The (easy) proof of the next result is omitted. The second part of the statement relies on Mezei's theorem characterizing the recognizable subsets of a direct product of monoids.

Proposition 4.1 *Let L be a language of A^*. If*

$$\pi_I(L) = \bigcup_{1 \leqslant i \leqslant n} L_{i,1} \times \cdots \times L_{i,k} \tag{5}$$

where for $1 \leqslant j \leqslant k$, the languages $L_{1,j}, \ldots, L_{n,j}$ are languages of A_j^*, then $[L]_I = \bigcup_{1 \leqslant i \leqslant n} L_{i,1}$ Ш $\cdots$ Ш $L_{i,k}$. In particular, if $\pi_I(L)$ is a recognizable subset of $A_1^* \times \cdots \times A_k^*$, then $[L]_I$ is regular.

If L is a group language, one can adapt an argument from [5, Proposition 9.6] to show that $\pi_I(L)$ can be decomposed as in (5), where each $L_{i,j}$ belongs to $\mathrm{Pol}(\mathcal{G})$. Therefore, since $\mathrm{Pol}(\mathcal{G})$ is closed under shuffle, we get:

Theorem 4.2 *Suppose that D is transitive. If L is a group language, then $[L]_I$ is a polynomial of group languages.*

Still some work is needed to obtain the following result.

Theorem 4.3 *Suppose that D is transitive. If L is a polynomial of group languages, then $[L]_I$ is also a polynomial of group languages.*

This result cannot be extended to $\mathcal{W}$. Indeed, let $A = \{a, b, c, d\}$ and $I = \{(a, b), (b, c), (c, d), (d, a)\}$. Then the language $(abcd)^* + A^*aaA^* + A^*bbA^* + A^*ccA^* + A^*ddA^* + A^*ababA^* + A^*bcbcA^* + A^*cdcdA^* + A^*dadaA^*$ belongs to $\mathcal{W}$ but $[L]_I$ is not regular, although D is transitive in this case.

4.2 The case where I is transitive

We now consider the case where I is transitive. In this case, $A^*/{\sim_I}$ is a free product of free commutative monoids.

Theorem 4.4 *Let L be a language of $\mathcal{W}(A^*)$ and let I be a transitive independence relation. Then $[L]_I$ is a regular language.*

Proof. (Sketch) Let $\mathcal{P} = \{A_1, \ldots, A_k\}$ be the partition of A such that $A^*/{\sim_I}$ is isomorphic to the free product $\mathbb{N}^{A_1} * \cdots * \mathbb{N}^{A_k}$.

Let $\mathcal{A} = (Q, A, \cdot, q_0, F)$ be the minimal automaton of L. Recall that the states of Q are partially ordered by the relation $\leqslant$ defined by $p \leqslant q$ if and only if, for all $u \in A^*$, $q\cdot u \in F$ implies $p\cdot u \in F$.

We now construct a generalized automaton $\mathcal{B}$, over the same set of states Q, in which transitions are labelled by regular languages. More precisely, for each pair of states (p, q), we create a transition from p to q labelled by

$$R_{p,q} = \bigcup_{1\leqslant i\leqslant k} \big[\{u \in A_i^* \mid p\cdot u \leqslant q\}\big]$$

Each language $\{u \in A_i^* \mid p\cdot u \leqslant q\}$ can be written as the intersection of A_i^* and of the language $K_{p,q} = \{u \in A^* \mid p\cdot u \leqslant q\}$. Since $L \in \mathcal{W}(A^*)$, one also has $K_{p,q} \in \mathcal{W}(A^*)$ and since $\mathcal{W}$ is closed under commutation by Theorem 3.5, so does $R_{p,q}$. The remainder of the proof (omitted for lack of space) consists in proving that $\mathcal{B}$ recognizes $[L]_I$. It follows that $[L]_I$ is regular. $\square$

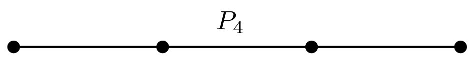
Note that we don't know whether $[L]_I$ also belongs to $\mathcal{W}(A^*)$. However, the proof of Theorem 4.4 can be adapted to prove another result.

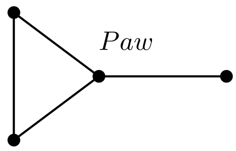
Let $I_1, \ldots, I_k$ be the connected components of the graph (A, I). Then $A^*/{\sim_I}$ is isomorphic to the free product $A^*/{\sim_{I_1}} * \cdots * A^*/{\sim_{I_k}}$. Let us modify the construction of the automaton $\mathcal{B}$ by taking

$$R_{p,q} = \bigcup_{1\leqslant j\leqslant k} \big[\{u \in A_j^* \mid p\cdot u \leqslant q\}\big]_{I_j}$$

Then one can prove that if each language $\left[\{u \in A_j^* \mid p\cdot u \leqslant q\}\right]_{I_j}$ is regular, then $[L]_I$ is regular. Putting $D_j = \{(a,b) \in A_j \times A_j \mid (a,b) \notin I_j\}$ for $1 \leqslant j \leqslant k$, one can show, thanks to Theorem 4.3, that $R_{p,q}$ is regular if L is a polynomial of group languages and each relation D_j is transitive.

There is a simple graph theoretic interpretation of this latter condition. One can show that I satisfies it if and only if the restriction of the graph (A, I) to any four letter subalphabet is not one of the graphs P_4 and Paw represented below:





We can now state our last result.

Theorem 4.5 *Let L be a polynomial of group languages. If the graph (A, I) is (P_4, Paw)-free, then $[L]_I$ is regular.*

References

[1] A. Achache, Opérateurs de fermeture semi-commutatifs, *Novi Sad J. Math.* **34**,1 (2004), 79–87.

[2] A. Bouajjani, A. Muscholl and T. Touili, Permutation Rewriting and Algorithmic Verification, in *Proc. 16th Symp. on Logic in Computer Science (LICS'01)*, Boston (MA), USA, 2001, pp. 399–409, IEEE Pub.

[3] A. Bouajjani, A. Muscholl and T. Touili, Permutation Rewriting and Algorithmic Verification, *Information and Computation* **205**,2 (2007), 199–224.

[4] W. Bucher, A. Ehrenfeucht and D. Haussler, On total regulators generated by derivation relations, *Theor. Comput. Sci.* **40**,2-3 (1985), 131–148.

[5] A. Cano Gómez and J.-E. Pin, On a conjecture of Schnoebelen, in *DLT 2003*, Z. Ésik and Z. Fülöp (ed.), Berlin, 2003, pp. 35–54, *Lect. Notes Comp. Sci.* n° 2710, Springer.

[6] A. Cano Gómez and J.-É. Pin, Shuffle on positive varieties of languages, *Theoret. Comput. Sci.* **312** (2004), 433–461.

[7] G. Cécé, P.-C. Héam and Y. Mainier, Efficiency of Automata in Semi-Commutation Verification Techniques, *Theoret. Informatics Appl.* **42** (2008), 197–215.

[8] A. De Luca and S. Varricchio, *Finiteness and regularity in semigroups and formal languages*, *Monographs in Theoretical Computer Science. An EATCS Series*, Springer-Verlag, Berlin, 1999.

[9] V. Diekert and Y. Métivier, Partial commutation and traces, in *Handbook of formal languages, vol. 3: beyond words*, pp. 457–533, Springer-Verlag New York, Inc., New York, NY, USA, 1997.

[10] V. Diekert and G. Rozenberg (ed.), *The book of traces*, World Scientific Publishing Co. Inc., River Edge, NJ, 1995.

[11] S. Ginsburg and E. H. Spanier, Bounded regular sets, *Proc. Amer. Math. Soc.* **17** (1966), 1043–1049.

[12] S. Ginsburg and E. H. Spanier, Semigroups, Presburger formulas, and languages, *Pacific J. Math.* **16** (1966), 285–296.

[13] P. Gohon, An algorithm to decide whether a rational subset of $\mathbb{N}^k$ is recognizable, *Theor. Comput. Sci.* **41** (1985), 51–59.

[14] G. Guaiana, A. Restivo and S. Salemi, On the product of trace languages, in *Proc. of the Workshop "Trace theory and code parallelization"*, A. Bertoni, M. Goldwurm and S. Crespi Reghizzi (ed.), University of Milan, Italy, 2000, pp. 54–67, Tech. Rep. n.263-00, 1-2 june 2000.

[15] G. Guaiana, A. Restivo and S. Salemi, On the trace product and some families of languages closed under partial commutations, *J. Autom. Lang. Comb.* **9**,1 (2004), 61–79.

[16] A. Muscholl and H. Petersen, A note on the commutative closure of star-free languages, *Inform. Process. Lett.* **57**,2 (1996), 71–74.

[17] E. Ochmański and K. Stawikowska, On closures of lexicographic star-free languages, in *Automata and formal languages*, pp. 227–234, Univ. Szeged. Inst. Inform., Szeged, 2005.

[18] E. Ochmański and K. Stawikowska, Star-free star and trace languages, *Fund. Inform.* **72**,1-3 (2006), 323–331.

[19] E. Ochmański and K. Stawikowska, A Star Operation for Star-Free Trace Languages, in *Developments in Language Theory, 11th International Conference, DLT 2007, Turku, Finland*, T. Harju, J. Karhumäki and A. Lepistö (ed.), pp. 337–345, *Lect. Notes Comp. Sci.* vol. 4588, Springer, 2007.

[20] J.-E. Pin, *Varieties of formal languages*, North Oxford, London et Plenum, New-York, 1986. (Traduction de Variétés de langages formels).

[21] J.-E. Pin, Polynomial closure of group languages and open sets of the Hall topology, *Theoret. Comput. Sci.* **169** (1996), 185–200.

[22] J. Sakarovitch, The "last" decision problem for rational trace languages, in *LATIN '92 (São Paulo, 1992)*, pp. 460–473, *Lecture Notes in Comput. Sci.* vol. 583, Springer, Berlin, 1992.