Closure of regular languages under semi-commutations*

Antonio Cano [1], Giovanna Guaiana[2] and Jean-Éric Pin[3]

Version 1.9 February 26, 2011 21 h 24

Abstract

The closure of a regular language under a [partial, semi-] commutation I has been extensively studied. We present new advances on two problems of this area: (1) When is the closure of a regular language under [partial, semi-] commutation still regular? (2) Are there any robust classes of languages closed under [partial, semi-] commutation? We show that the class $\mathrm{Pol}(\mathcal{G})$ of polynomials of group languages is closed under commutation, and under partial commutation when the complement of I in A^2 is a transitive relation. We also give a sufficient graph theoretic condition on I to ensure that the closure of a language of $\mathrm{Pol}(\mathcal{G})$ under I-commutation is regular. We exhibit a very robust class of languages $\mathcal{W}$ which is closed under commutation. This class contains $\mathrm{Pol}(\mathcal{G})$ and is decidable. It is also closed under intersection, union, shuffle, concatenation, quotients, length-decreasing morphisms and inverses of morphisms. If I is transitive, we show that the closure of a language of $\mathcal{W}$ under I-commutation is regular. Finally, we prove a few results on semi-commutations. The proofs are nontrivial and combine several advanced techniques, including combinatorial Ramsey type arguments, algebraic properties of the syntactic monoid, finiteness conditions on semigroups and properties of insertion systems.

The closure of a regular language under commutation, partial commutation or semi-commutation has been extensively studied [35, 24, 1, 16, 17, 18], notably in connection with regular model checking [2, 3, 8, 9] or in the study of Mazurkiewicz traces, one of the models of parallelism [20, 21, 25, 36]. We refer the reader to the survey [15, 14] or to the recent articles of Ochmański [26, 27, 28] 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

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

[2]LITIS EA 4108, Université de Rouen, BP12, 76801 Saint Etienne du Rouvray, France.

[3]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. The first author was supported by the project *Técnicas de Inferencia Gramatical y aplicación al procesamiento de biosecuencias* (TIN2007-60769) supported by the Spanish Ministery of Education and Sciences. The third author was supported by the project ANR 2010 BLAN 0202 02 FREC.

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, quotients, 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 operations, quotients 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, semi-*]*commutation still regular?*

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

Apart from group languages, the classes considered in this paper are all closed under polynomial operations. 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, length-preserving morphisms and inverses of morphisms. There is also a very robust class of languages, denoted $\mathcal{W}$, which contains $\mathrm{Pol}(\mathcal{G})$ and is closed under intersection, union, shuffle, concatenation, quotients, length-decreasing morphisms and inverses of morphisms [7]. This class is decidable and can be defined as the largest positive variety of languages not containing $(ab)^*$.

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:

(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) The class $\mathcal{W}$ is closed under commutation.

(4) If I is transitive, the closure of a language of $\mathcal{W}$ under I is regular.

(5) If I is the semi-commutation $ab \to ba$ on the alphabet $\{a, b\}$, the closure of a language of $\mathrm{Pol}(\mathcal{G})$ under I is also in $\mathrm{Pol}(\mathcal{G})$.

(6) Let L be a group language of A^* and let n be the size of its syntactic monoid. For each semi-commutation I on A, there exists an integer k such that, for any letter $a \in A$, $a^k \sim_{[L]_I} a^{k+n}$.

Result (3) is probably the most important of these 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 [6, 7], finiteness conditions on semigroups [13] and properties of insertion systems [4]. A part of these results were first presented in [5].

Our paper is organised as follows. We first survey the known results in Section 2. Then we establish some combinatorial properties, notably on group languages in Section 3. Our results on commutative closure are established in Section 4 and those on closure under partial commutation in Section 5. Section 6 is devoted to semi-commutations.

1 Definitions and notation

1.1 Words and subwords

In this paper, A denotes a finite alphabet and A^* is the free monoid on A. The empty word is denoted by 1. A word u is a *subword* of v if v can be written as

$$v = v_0u_1v_1u_2v_2 \cdots u_kv_k$$

where u_i and v_i are words (possibly empty) such that $u_1u_2 \cdots u_k = u$. For instance, the words *baba* and *acab* are subwords of *abcacbab*.

1.2 Commutations and semi-commutations

Let A be an alphabet. A *semi-commutation* is an irreflexive relation on A. A *partial commutation* is a symmetric and irreflexive relation on A, often called the *independence relation* in the literature. To each semi-commutation I is associated the rewriting system $\{ab \to ba \mid (a, b) \in I\}$. Given two words $u, v \in A^*$, we write $u \to_I v$ if there is a pair $(a, b) \in I$ and two words p and s such that $u = pabs$ and $v = pbas$. We denote by $\xrightarrow{*}_I$ the reflexive transitive closure of $\to_I$. If I is a partial commutation, then $\xrightarrow{*}_I$ is a congruence on A^* and for this reason, the notation $\sim_I$ is preferred.

If L is a language on A^*, we denote by $[L]_I$ the closure of L under $\xrightarrow{*}_I$. More precisely,

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

A class $\mathcal{C}$ of languages is *closed under I-semi-commutation* if $L \in \mathcal{C}$ implies $[L]_I \in \mathcal{C}$. 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. A class of languages $\mathcal{C}$ is *closed under commutation* if $L \in \mathcal{C}$ implies $[L] \in \mathcal{C}$.

The *non-commutation 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. These graphs are undirected in the case of a (partial) commutation and directed for a semi-commutation.

1.3 Operations on languages

The *marked product* of $k+1$ languages L_0, L_1, ..., L_k of A^* is a product of the form $L = L_0a_1L_1 \cdots a_kL_k$, where $a_1, \ldots, a_k$ are letters of A.

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

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

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

Given a class $\mathcal{L}$ of regular languages, the *polynomial closure* of $\mathcal{L}$, denoted by $\mathrm{Pol}(\mathcal{L})$, consists of 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 $\mathcal{L}$. For instance, if $\mathcal{I}$ is the trivial class of languages defined by $\mathcal{I}(A^*) = \{\emptyset, A^*\}$ for each alphabet A, then $\mathrm{Pol}(\mathcal{I})$ is the class of finite unions of languages of the form $A^*a_1A^* \cdots a_kA^*$, with $a_1, \ldots, a_k \in A$.

A *morphism* between two free monoids A^* and B^* is a map $\varphi : A^* \to B^*$ such that, for all $u, v \in A^*$, $\varphi(uv) = \varphi(u)\varphi(v)$. This condition implies in particular that $\varphi(1) = 1$. We say that φ is *length-preserving* if, for each $u \in A^*$, the words u and $\varphi(u)$ have the same length. Equivalently, φ is length-preserving if, for each letter $a \in A$, $\varphi(a) \in B$. Similarly, φ is *length-decreasing* if the image of each letter is either a letter or the empty word.

1.4 Syntactic ordered monoid

Let L be a regular language of A^*. The *syntactic preorder* of L is the relation $\leqslant_L$ defined on A^* by : $u \leqslant_L v$ iff, for every $x, y \in A^*$,

$$xvy \in L \Rightarrow xuy \in L$$

The *syntactic congruence* of L is the relation $\sim_L$ defined on A^* by : $u \sim_L v$ iff, for every $x, y \in A^*$,

$$xvy \in L \Leftrightarrow xuy \in L$$

The *syntactic ordered monoid* of L is $(A^*/{\sim_L}, \leqslant_L /{\sim_L})$, where $\leqslant_L /{\sim_L}$ denotes the order induced by $\leqslant_L$ on the quotient set $A^*/{\sim_L}$.

The syntactic ordered monoid can be computed from the minimal automaton as follows. First observe that if $\mathcal{A} = (Q, A, \cdot, q_-, F)$ is a minimal deterministic automaton, the relation $\leqslant$ defined on Q by $p \leqslant q$ if for all $u \in A^*$,

$$q \cdot u \in F \Rightarrow p \cdot u \in F$$

is an order relation, called the *syntactic order*. Then the syntactic ordered monoid of a language is the transition monoid of its ordered minimal automaton. The order is defined by $u \leqslant v$ if and only if, for all $q \in Q$, $q \cdot u \leqslant q \cdot v$.

Example 1.1 The minimal deterministic automaton of $(ab)^*$ is represented in Figure 1.1.

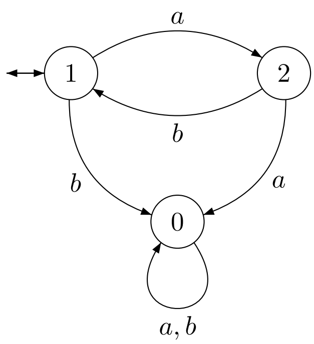


Figure 1.1: The minimal deterministic automaton of $(ab)^*$

The order on the set of states is $1 < 0$ and $2 < 0$. Indeed, one has $0 \cdot u = 0$ for all $u \in A^*$ and thus, the formal implication

$$0 \cdot u \in F \Rightarrow q \cdot u \in F$$

holds for any state q. One can verify that there is no other relations among the states. For instance, 1 and 2 are incomparable since $1 \cdot ab = 1 \in F$ but $2 \cdot ab = 0 \notin F$ and $1 \cdot b = 0 \notin F$ but $2 \cdot b = 1 \in F$.

The syntactic monoid of $(ab)^*$ and its syntactic order are represented below :

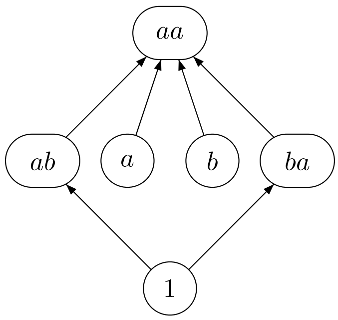
Elements

	1	2
1	1	2
a	2	0
b	0	1
aa	0	0
ab	1	0
ba	0	2

Relations

$$bb = aa = 0$$
$$aba = a$$
$$bab = b$$

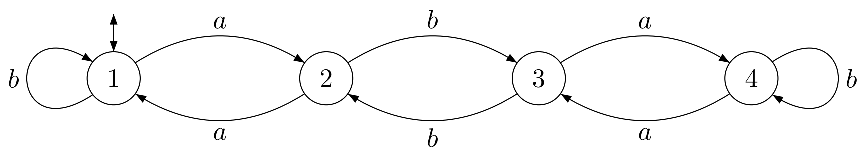
Syntactic order



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. The definition of the star-free languages follows the same definition scheme as the one of rational languages, with the difference that the star operation is replaced by the complement. Thus the *star-free* languages of A^* are obtained from the finite languages by using Boolean operations and concatenation product. A well-known result of Schützenberger states that a regular language is star-free if and only if its syntactic monoid has period 1.

Opposite to the star-free languages are the group languages. Recall that a *group language* is a language whose syntactic monoid is a group, or, equivalently, is recognised by a finite deterministic automaton in which each letter defines a permutation of the set of states. Note that if a group language is recognised by a group G, then its period divides $|G|$.

Example 1.2 The set of words over $A = \{a, b\}$ having an even number of subwords equal to ab is a group language whose syntactic monoid is the group D_4. A regular expression for this language is $\big(b + a(b(ab^*a)^*b)^*a\big)^*$ and its minimal automaton is represented below.



2 Known results

In this section, we briefly survey the kwown results on our two problems.

2.1 The first problem

For the commutative closure, the problem is solved [35, 16, 17, 18]:

Theorem 2.1 *One can decide whether the commutative closure of a given regular language is regular.*

The commutative closure of the language $(ab)^*$ is not regular since $[(ab)^*] = \{u \in \{a,b\}^* \mid |u|_a = |u|_b\}$. Unfortunately, the class of languages whose commutative closure is regular is not robust. In particular, it is not even closed under intersection as shown in the next example.

Example 2.1 Consider the languages $L_1 = (ab)^* + (ab)^* a^+ b^+$ and $L_2 = (ab)^* + (ab)^* b^+ a^+$. The commutative closure of these languages is regular, since

$$[L_1] = [L_2] = \{a,b\}^* \setminus (a^+ + b^+)$$

However, $L_1 \cap L_2 = (ab)^*$ and $[(ab)^*]$ is not regular.

For partial commutations, the result of Sakarovitch [36] concluded a series of previous partial results.

Theorem 2.2 *One can decide whether the closure $[L]_I$ of a regular language L is regular if and only if I is a transitive relation.*

For semi-commutations, the following useful result holds [10].

Theorem 2.3 *Let I be a semi-commutation on A and let $L_1, \ldots, L_n$ be languages of A^*. If the languages $[L_1]_I, \ldots, [L_n]_I$ are regular, then $[L_1 \cdots L_n]_I$ is regular.*

2.2 The second problem

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

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

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

(3) the class $\text{Pol}(\mathcal{J})$, which consists of finite unions of languages of the form $A_0^* a_1 A_1^* \cdots a_k A_k^*$ with $A_i \subseteq A$ and $a_1, \ldots, a_k \in A$, also called *APC* (*Alphabetic Pattern Constraints*) in [2],

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

Syntactic characterizations are known for $\mathcal{J}$ [37] and for $\text{Pol}(\mathcal{J})$ [33]. The following theorem summarises the results of Guaiana, Restivo and Salemi [20, 21], Bouajjani, Muscholl and Touili [2, 3] and Cécé, Héam and Mainier [8, 9].

Theorem 2.4 *The following properties hold:*

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

Note that neither Pol($\mathcal{I}$) nor $\mathcal{J}$ are closed under partial commutation [21, Theorem 15].

2.3 Star-free languages

Two nice results on star-free languages were proved by Muscholl and Petersen [25]. The first one is the counterpart of Theorem 2.2 for star-free languages.

Theorem 2.5 *Let I be a partial commutation. 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 2.6 *Let I be a partial commutation and 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.*

Let us remind the example given in [25]. The language $(abcbac)^*$ is star-free, whereas the language $[L]_{ab=ba} = (((ab + ba)c)^2)^*$ is regular but not star-free.

3 Some combinatorial properties

In this section, we gather together the combinatorial properties that are used in this paper. We first state some consequences of Ramsey's theorem, then we prove some properties of group languages. Finally, we establish a few results on insertion systems.

3.1 Ramsey type properties

In this section, we briefly survey a few consequences of a celebrated result in combinatorics on words, Ramsey's theorem. Similar results can be found for instance in [13, 22, 30], with a slightly different formulation. For this reason, we prefer to give here a self-contained proof.

Let us first state Ramsey's theorem. If E is a finite set, we denote by $|E|$ the number of elements of E. If $|E| = n$, then E is called an *n-set.* A *colouring* of a set E in m colours is a function from E into $\{1, \ldots, m\}$.

Theorem 3.1 (Ramsey) *Let r, k and m be positive integers such that $k \geqslant r$. Then there exists an integer $N = R(r, k, m)$ such that, for each N-set E and for each colouring in m colours of the r-subsets of E, there exists a k-subset of E whose r-subsets have the same colour.*

The next proposition is a consequence of Ramsey's theorem.

Proposition 3.2 *Let M be a finite monoid and let $\pi : A^* \to M$ be a surjective morphism. For any $n > 0$, there exists $N > 0$ and an idempotent e in M 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 $\pi(u_{i_0} u_{i_0+1} \cdots u_{i_1-1}) = \pi(u_{i_1} u_{i_1+1} \cdots u_{i_2-1}) = \ldots = \pi(u_{i_{n-1}} \cdots u_{i_n-1}) = e$.*

Proof. Let $N = R(2, n+1, |M|)$ and let $u_0, u_1, \ldots, u_N$ be words of A^*. Let $E = \{0, \ldots, N\}$. We define a colouring into $|M|$ colours of the 2-subsets of E in the following way: the colour of the 2-subset $\{i, j\}$ (with $i < j$) is the element $\pi(u_i u_{i+1} \cdots u_{j-1})$ of M. According to Ramsey's theorem, one can find $n+1$ indices $i_0 < i_1 < \cdots < i_n$ such that all the 2-subsets of $\{i_0, \ldots, i_n\}$ have the same colour. In particular, since $n + 1 \geqslant 2$, one gets

$$\begin{aligned}\pi(u_{i_0} u_{i_0+1} \cdots u_{i_1-1}) &= \pi(u_{i_1} u_{i_1+1} \cdots u_{i_2-1}) = \ldots = \pi(u_{i_{n-1}} \cdots u_{i_n-1}) \\ &= \pi(u_{i_0} u_{i_0+1} \cdots u_{i_2-1})\end{aligned}$$

Let e be the common value of these elements. It follows from the equalities $\pi(u_{i_0} u_{i_0+1} \cdots u_{i_1-1}) = \pi(u_{i_1} u_{i_1+1} \cdots u_{i_2-1}) = \pi(u_{i_0} u_{i_0+1} \cdots u_{i_2-1})$ that $ee = e$ and thus e is idempotent. $\square$

When M is a finite group, 1 is the unique idempotent of M and Proposition 3.2 can be simplified as follows:

Corollary 3.3 *Let G be a finite group and let $\pi : A^* \to G$ be a surjective morphism. Then 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 $\pi(u_{i_0} u_{i_0+1} \cdots u_{i_1-1}) = \pi(u_{i_1} u_{i_1+1} \cdots u_{i_2-1}) = \ldots = \pi(u_{i_{n-1}} \cdots u_{i_n-1}) = 1$.*

3.2 Properties of group languages

In this section, we establish some simple properties of group languages. Let us start with an elementary lemma.

Lemma 3.4 *Let $g_1, g_2, \ldots, g_{|G|}$ be a sequence of elements of G. Then there exist two indices i, j, with $i \leqslant j \leqslant |G|$ such that $g_i \cdots g_j = 1$.*

Proof. Consider the sequence $g_1, g_1 g_2, \ldots, g_1 g_2 \cdots g_{|G|}$. Either one of these elements is equal to 1, or two of them are equal, say $g_1 \cdots g_{i-1} = g_1 \cdots g_j$ with $i \leqslant j$. In this case, $g_i \cdots g_j = 1$. $\square$

The next lemma is a kind of insertion property. Let π be a morphism from A^* onto a finite group G, let $R = \pi^{-1}(1)$ and let L be a language recognised by π.

Lemma 3.5 *Let x be a word of R and let u and v be two words. Then $uv \in L$ if and only if $uxv \in L$.*

Proof. If $x \in R$, then $\pi(x) = 1$. It follows that

$$\pi(uxv) = \pi(u)\pi(x)\pi(v) = \pi(u)\pi(v) = \pi(uv)$$

which proves the lemma. $\square$

We shall also need the following consequence of the previous lemma.

Lemma 3.6 *Let $a_1, \ldots, a_r$ be letters, let x be a word of R and let u and v be two words. If $uv \in Ra_1Ra_2R \cdots Ra_rR$, then $uxv \in Ra_1Ra_2R \cdots Ra_rR$.*

Proof. If $uv \in Ra_1Ra_2R \cdots Ra_rR$, then there exist an index i and two words $x', x'' \in A^*$ such that $u \in Ra_1R \cdots Ra_ix'$, $v \in x''a_{i+1}R \cdots Ra_rR$ and $x'x'' \in R$. Since $x'xx'' \in R$ by Lemma 3.5, one gets $uxv \in Ra_1Ra_2R \cdots Ra_rR$. □

3.3 Insertion systems

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 $\to_R$. The closures of a language L of A^* under $\to_R$ and $\xrightarrow{*}_R$ are respectively the languages

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

We are especially interested in the case $R = \pi^{-1}(1)$, where π is a morphism from A^* onto a finite group G. In this case, the set of words that can be derived from a given word has a simple expression. Let us introduce a convenient (but nonstandard!) notation to state this result more easily. Given a word $u = a_1 \cdots a_n$ and a language K, let us denote by $u \uparrow K$ the language $Ka_1K \cdots Ka_nK$.

Proposition 3.7 *For each word u of A^*, one has $[u]_{\xrightarrow{*}_R} = u \uparrow R$.*

Proof. The inclusion of $u \uparrow R$ in $[u]_{\xrightarrow{*}_R}$ is an immediate consequence of the definitions. For the opposite inclusion, since $u \in u \uparrow R$, it suffices to prove that the language $u \uparrow R$ is closed under $\to_R$. But this is just another formulation of Lemma 3.6. □

Let F be the set of words of R of length $\leqslant |G|$. Then F is finite by construction. The next lemma states that sufficiently long words contain a factor in F.

Lemma 3.8 *Every word of A^* of length $\geqslant |G|$ contains a nonempty factor in F.*

Proof. Let $a_1 \cdots a_n$ be a word of length $n \geqslant |G|$. By Lemma 3.4, there exist two indices i, j, with $i \leqslant j \leqslant |G|$ such that $\pi(a_i) \cdots \pi(a_j) = 1$. It follows that $\pi(a_i \cdots a_j) = 1$ and hence $a_i \cdots a_j \in F$. □

The following result can be viewed as a special case of a well-known result [23, Proposition I.6.4].

Proposition 3.9 *The relations $\xrightarrow{*}_F$ and $\xrightarrow{*}_R$ coincide.*

Proof. Since $F \subseteq R$, it is clear that $u \xrightarrow{*}_F v$ implies $u \xrightarrow{*}_R v$. Since $\xrightarrow{*}_F$ is transitive, it is now sufficient to show that $u \to_R v$ implies $u \xrightarrow{*}_F v$. Thus suppose that $u = u'u''$, and $v = u'ru''$ for some $r \in R$. We prove the result by induction on the length of r. If $|r| \leqslant |G|$, then $r \in F$ and $u \to_F v$. Otherwise, Lemma 3.8 shows that r contains a nonempty factor in F. Thus $r = xfy$ with $f \in F$. Further, Lemma 3.5 shows that $xy \in R$. Thus $u \to_R u'xyu''$ and by the induction hypothesis, $u \xrightarrow{*}_F u'xyu''$. Now, since $u'xyu'' \to_F u'xfyu'' = v$, one has $u \xrightarrow{*}_F v$. $\square$

Let us prove a key property of $\xrightarrow{*}_R$. Recall that a *well quasi-order* on a set E is a reflexive and transitive relation $\leqslant$ such that for any infinite sequence $x_0, x_1, \ldots$ of elements of E, there are two integers $i < j$ such that $x_i \leqslant x_j$.

Proposition 3.10 *The relation $\xrightarrow{*}_R$ is a well quasi-order on A^*.*

Proof. The proof relies on a powerful result of [4]: if H is a finite set of words such that the language $A^* \setminus A^*HA^*$ is finite, then the relation $\xrightarrow{*}_H$ is a well quasi-order on A^*. Lemma 3.8 shows that the set F satisfies these conditions and thus $\xrightarrow{*}_F$ is a well quasi-order on A^*. Further, Proposition 3.9 shows that $\xrightarrow{*}_R$ is equal to $\xrightarrow{*}_F$. $\square$

We now derive an important consequence of Proposition 3.10.

Proposition 3.11 *For each language L of A^*, the language $[L]_{\xrightarrow{*}_R}$ is a polynomial of group languages.*

Proof. Since $\xrightarrow{*}_R$ is a well quasi-order, the language $[L]_{\xrightarrow{*}_R}$ is equal to $[G]_{\xrightarrow{*}_R}$ for some finite language G. It follows that $[L]_{\xrightarrow{*}_R}$ is a finite union of languages of the form $[u]_{\xrightarrow{*}_R}$. It follows from Proposition 3.7 that $[L]_{\xrightarrow{*}_R}$ is a polynomial of group languages. $\square$

Corollary 3.12 *A language L that satisfies $L = [L]_{\to_R}$ is a polynomial of group languages.*

Proof. Indeed, the equality $L = [L]_{\to_R}$ implies $L = [L]_{\xrightarrow{*}_R}$ and by Proposition 3.11, the language $[L]_{\xrightarrow{*}_R}$ is a polynomial of group languages. $\square$

4 Commutative closure

This section contains three series of results. The first one concerns group languages, the second one polynomials of group languages and the third one a robust class of languages introduced in [6, 7] and denoted by $\mathcal{W}$.

4.1 Group languages

The main result of this section states that the commutative closure of a group language is regular, and is in fact a polynomial of group languages. We start with a proof of the weaker property, which relies only on Ramsey type arguments and will serve as a guide for the more technical proof of Theorem 4.6.

Theorem 4.1 *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 Corollary 3.3. We claim that for any letter $a \in A$, $a^N \sim_{[L]} a^{N+n}$. Let $g = \pi(a)$.

Suppose that $xa^Ny \in [L]$. Then there exists a word w of L commutatively equivalent to xa^Ny. 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 Corollary 3.3 to the sequence of words u_0a, u_1a, …, 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

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

Let r and s be the words defined by

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

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

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

is commutatively equivalent to xa^Ny. 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 $xa^Ny \in [L]$, 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. □

Theorem 4.1 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 4.1.

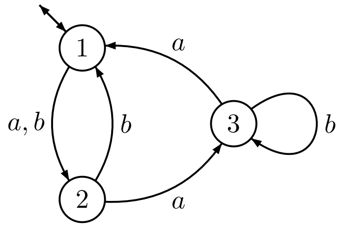
Theorem 4.2 *The commutative closure of a group language is a polynomial of group languages.*

Proof. Let L be a group language, let $\pi : A^* \to G$ be its syntactic morphism and let $R = \pi^{-1}(1)$. Let K be the commutative closure of L. We claim that $K = [K]_{\to R}$. It suffices to prove that if $xy \in K$ and $r \in R$, then $xry \in K$. Since $xy \in K$, there exists a word $v \in L$ which is commutatively equivalent to xy. Thus the word vr is commutatively equivalent to xry. Now since $\pi(r) = 1$, one gets

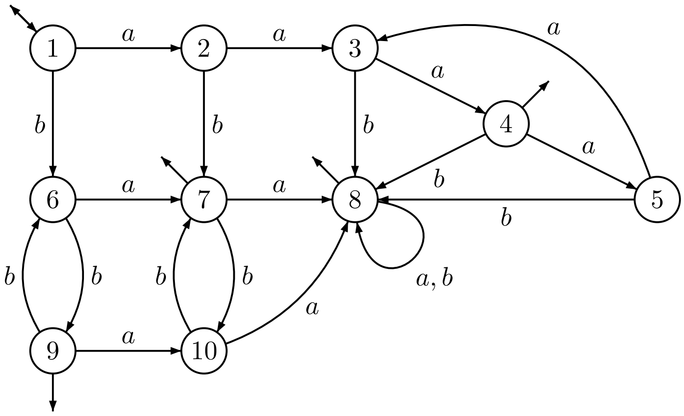
$$\pi(vr) = \pi(v)\pi(r) = \pi(v)$$

Therefore $vr \in L$ and $xry \in K$, which proves the claim. It follows by Corollary 3.12 that K is a polynomial of group languages. □

Example 4.1 Let $A = \{a, b\}$ and let L be the group language of A^* accepted by the automaton represented below.



Thus L is recognised by the group of all permutations of a three-element set. Its commutative closure is the language $L_1 + (a^3)^* + (b^2)^* + (b^2)^*ab(b^2)^* + (b^2)^*ba(b^2)^*$, where $L_1 = A^*aA^*aA^*bA^* + A^*aA^*bA^*aA^* + A^*bA^*aA^*aA^*$. Its minimal automaton is the following



Finally, one can write $[L]$ as a polynomial of group languages as follows: $[L] = L_1 + L_2$ where L_2 is the group language defined by

$$L_2 = \{u \in A^* \mid |u|_a \equiv 0 \bmod 3 \text{ and } |u|_b \equiv 0 \bmod 2 \text{ or } |u|_a \equiv 1 \bmod 3 \text{ and } |u|_b \equiv 1 \bmod 2\}.$$

The next example shows that the commutative closure of a group language is not in general a group language.

Example 4.2 Let L be the set of words over $A = \{a, b\}$ having an odd number of subwords equal to ab. Then L is a group language, but its commutative closure $A^*aA^*bA^* \cup A^*bA^*aA^*$ is not a group language.

4.2 Polynomials of group languages

Let us first recall some basic facts about polynomial of group languages.

Theorem 4.3 *The class Pol($\mathcal{G}$) is closed under shuffle, product and marked product.*

Proof. It was shown in [34] that the positive variety of languages Pol($\mathcal{G}$) corresponds to the variety of finite ordered monoids $\mathbf{PG}^+$. It follows then from the results of [7] that Pol($\mathcal{G}$) is closed under shuffle. It is also closed under marked product by construction.

Let L and L' be two languages. Then

$$LL' = \begin{cases} \bigcup_{a \in A} La(a^{-1}L') & \text{if } 1 \notin L' \\ \bigcup_{a \in A} La(a^{-1}L') \cup L & \text{if } 1 \in L' \end{cases}$$

Let now L and L' be two group languages. Since group languages are closed under quotients, $a^{-1}L'$ is a group language. It follows that LL' belongs to Pol($\mathcal{G}$) and it follows immediately that Pol($\mathcal{G}$) is closed under product. □

We now extend Theorem 4.2 as follows.

Theorem 4.4 *The commutative closure of a polynomial of group languages is also a polynomial of group languages.*

Proof. It is shown in [32] 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 monomials of the form $Ra_1R \cdots Ra_nR$, where $R = \pi^{-1}(1)$ and $a_1, \ldots, a_n$ are letters of A. Clearly, it suffices to prove the theorem when L is one of these monomials. Let K be its commutative closure. By Corollary 3.12, it suffices to prove that $K = [K]_{\to R}$ to show that K is a polynomial of group languages.

Let x, y and r be words such that $xy \in K$ and $r \in R$. Let v be a word of L commutatively equivalent to xy. Then vr is commutatively equivalent to xry. As an element of L, v can be written as $r_0a_1r_1 \cdots a_nr_n$ for some words $r_0, \ldots, r_n \in R$. Thus $vr \in L$ since $r_nr \in R$. It follows that $xry \in K$ and hence $K = [K]_{\to R}$. □

4.3 Languages of $\mathcal{W}$

We now define the class of regular languages $\mathcal{W}$ first introduced and studied in [6, 7]. Recall that a *positive variety* of languages is a class of regular languages closed under union, intersection, quotients 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-decreasing morphisms and inverses of morphisms, $\mathcal{W}$ is a quite robust class, which strictly contains the classes APC, Pol($\mathcal{C}om$) and Pol($\mathcal{G}$).

The class $\mathcal{W}$ has an algebraic characterization [6, 7]. 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 $(M, \leqslant)$ satisfies the following condition $(*)$:

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 [7, p. 435–436] for a precise definition of the semigroup notions used in $(*)$. The finite ordered monoids satisfying $(*)$ form a variety of ordered monoids $\mathbf{W}$. Condition $(*)$ 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 relates the period of a language of $\mathcal{W}$ to the period of its commutative closure. We will need the following proposition.

Proposition 4.5 *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 [13, 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. $\square$

The main result of this section can now be stated.

Theorem 4.6 *Let L be a language of $\mathcal{W}(A^*)$. Then $[L]$ belongs to $\mathcal{W}(A^*)$ and its period divides that of L.*

Proof. Let L be a language of $\mathcal{W}(A^*)$ and let $[L]$ be its commutative closure. Since $[L]$ is commutative and since $\mathcal{W}$ contains the variety of commutative languages, proving that $[L]$ belongs to $\mathcal{W}(A^*)$ amounts to show that $[L]$ is regular.

Since $L \in \mathcal{W}(A^*)$, there exist an ordered monoid $(M, \leqslant) \in \mathbf{W}$, a surjective monoid morphism $\pi : A^* \to M$ and an order ideal P of $(M, \leqslant)$ such that $\pi^{-1}(P) = L$. Let ω, p and n be respectively the exponent, the period and the size of M. 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 4.5 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 three combinatorial results. The first one is almost trivial.

Proposition 4.7 *For every $m \in M$, there exists a word u of length $\leqslant n$ such that $\pi(u) = m$.*

Proof. Let $m \in M$ and let $u = a_1 \cdots a_{|u|}$ be a word of minimal length in $\pi^{-1}(m)$. Suppose that $|u| \geqslant n$. Then, by the pigeonhole principle, two of the $n+1$ elements $\pi(1), \pi(a_1), \pi(a_1a_2), \ldots, \pi(a_1 \cdots a_n)$ are equal, say $\pi(a_1 \cdots a_i)$ and $\pi(a_1 \cdots a_j)$ with $i < j$. It follows that $\pi(u) = \pi(a_1 \cdots a_i a_{j+1} \cdots a_{|u|})$, which contradicts the definition of u. Thus $|u| \leqslant n$. □

The second one is a slight variation of Proposition 3.2.

Proposition 4.8 *Let c be a letter of an alphabet A. For any $r > 0$, there exists an integer $N = N(r)$ 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 $u = v_0v_1cv_2c \cdots v_rcv_{r+1}$ such that, for $1 \leqslant i \leqslant r$, $\pi(v_ic) = e$.*

Proof. Let u be a word containing at least $N+1$ occurrences of c. Let us write this word as $u = u_0cu_1c \cdots u_Ncu_{N+1}$, where, for $0 \leqslant i \leqslant N+1$, $u_i \in A^*$. By Proposition 3.2, applied to the words $u_0c, \ldots, u_Nc$, there exist integers $0 \leqslant i_0 < i_1 < \ldots < i_r \leqslant N$ and an idempotent e of M such that

$$\pi(u_{i_0}c \cdots u_{i_1-1}c) = \ldots = \pi(u_{i_{r-1}}c \cdots u_{i_r-1}c) = e$$

Setting

$$\begin{aligned} v_0 &= u_0c \cdots u_{i_0-1}c \\ v_1 &= u_{i_0}c \cdots u_{i_1-2}cu_{i_1-1} \\ &\vdots \\ v_r &= u_{i_{r-1}}c \cdots u_{i_r-2}cu_{i_r-1} \\ v_{r+1} &= u_{i_r}c \cdots u_Ncu_{N+1} \end{aligned}$$

we obtain a factorization $u = v_0v_1c \cdots v_rcv_{r+1}$ such that, for $1 \leqslant i \leqslant r$, $\pi(v_ic) = e$. □

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

Proposition 4.9 *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 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 $\gamma(B^*)$. By Proposition 4.7, applied to γ, there exists 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 $\gamma(z) \in J$. □

Let z be the balanced word given by Proposition 4.9. 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 Proposition 4.8.

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 Proposition 4.8, there exist an idempotent e of M and a factorization

$$u = v_0 v_1 c \cdots v_{n_1} c v_{n_1+1}$$

such that, for $1 \leqslant i \leqslant n_1$, $\pi(v_i c) = e$.

Now, since $n_1 = 3n_2$, one can also write u as

$$u = v_0 (f_1 g_1) \cdots (f_{n_2} g_{n_2}) v_{n_1+1}$$

where, for $1 \leqslant i \leqslant n_2$, $f_i = v_{3i-2} c v_{3i-1}$ and $g_i = c v_{3i} c$. The next lemma is the key argument to the proof of Theorem 4.6.

Lemma 4.10 *For $1 \leqslant i \leqslant n_2$, the elements $\pi(f_i)$ and $\pi(g_i)$ are mutually inverse.*

Proof. The result follows from the following formulas:

$$\begin{aligned}
&\pi(f_i)\pi(g_i)\pi(f_i)\\
&\qquad = \pi(v_{3i-2}c)\pi(v_{3i-1}c)\pi(v_{3i}c)\pi(v_{3i-2}c)\pi(v_{3i-1})\\
&\qquad = e\pi(v_{3i-1}) = \pi(v_{3i-2}c)\pi(v_{3i-1}) = \pi(f_i)\\
&\pi(g_i)\pi(f_i)\pi(g_i)\\
&\qquad = \pi(c)\pi(v_{3i}c)\pi(v_{3i-2}c)\pi(v_{3i-1}c)\pi(v_{3i}c)\\
&\qquad = \pi(c)e = \pi(c)\pi(v_{3i}c) = \pi(g_i)
\end{aligned}$$

□

Setting $\bar{s} = \pi(c)e$, one gets $\pi(g_i) = \bar{s}$ 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 $\pi(f_{i_1}) = \ldots = \pi(f_{i_{n_3}}) = s$. Setting

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

we obtain a factorization

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

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

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 $y_{d+(j-1)r+2}$, ..., the r's occurrence of b by y_{d+jr}. Finally, set

$$u' = w_0(v_{3i_1-2}ccv_{3i_1-1}cz_1v_{3i_1}c)(v_{3i_2-2}ccv_{3i_2-1}cz_2v_{3i_2}c)\ \cdots \\ (v_{3i_d-2}ccv_{3i_d-1}cz_dv_{3i_d}c)w_1\ \cdots\ w_{n_3} \quad (4)$$

We are now ready for the three final steps.

Lemma 4.11 *The word u' is commutatively equivalent to $xc^{N+d}y$.*

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

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

Now,

$$v_{3i_1-2}cv_{3i_1-1}cv_{3i_1}c = f_{i_1}g_{i_1} = x_1y_1$$
$$\vdots$$
$$v_{3i_d-2}cv_{3i_d-1}cv_{3i_d}c = f_{i_d}g_{i_d} = x_dy_d$$

Further, by construction, $z_1\ \cdots\ z_d \sim x_{d+1}y_{d+1}\ \cdots\ x_{n_3}y_{n_3}$. Therefore

$$u' \sim c^d w_0x_1y_1w_1x_2y_2w_2\ \cdots\ w_{n_3-1}x_{n_3}y_{n_3}w_{n_3}$$

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 4.9, $\gamma(z)$ 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 4.12 *One has $\pi(z_1) = \ldots = \pi(z_d) = \gamma(z)$.*

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 π, namely s (resp. $\bar{s}$), $\pi(z_j)$ is equal to $\gamma(z)$. $\square$

Lemma 4.13 *The word u' belongs to L.*

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

$$\begin{aligned}\pi(v_{3i_1-2}ccv_{3i_1-1}cz_1v_{3i_1}c) &= \pi(v_{3i_1-2}c)\pi(c)\pi(v_{3i_1-1}c)\pi(z_1)\pi(v_{3i_1}c)\\ &= e\pi(c)e\pi(z_1)e = e\bar{s}\gamma(z)e \quad \text{by Lemma 4.12}\end{aligned}$$

By a similar argument, one has

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

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

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

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 $\pi(L)$ is an order ideal, the element $\pi(w_0)(e\bar{s}\gamma(z)e)^d\pi(w_{n_3})$ is also in $\pi(L)$ and hence $u' \in L$. □

Putting Lemmas 4.11 and 4.13 together, we conclude that $xc^{N+d}y \in [L]$, which proves the claim and the theorem. □

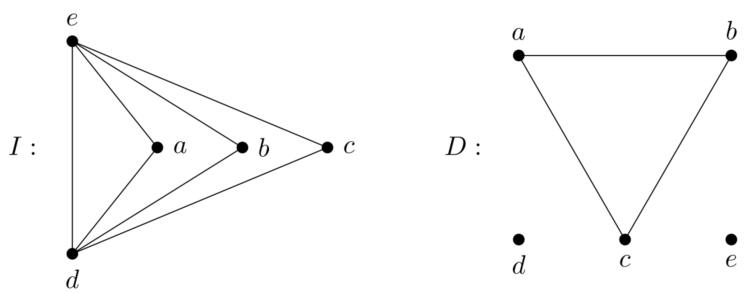
Note that there are regular languages outside of $\mathcal{W}$ whose commutative closure is in $\mathcal{W}$. For instance the language $(ab)^*(a^*+b^*)$ is not in $\mathcal{W}$ but its commutative closure is A^*.

5 Closure under partial commutation

Some of the results of Section 4 can be extended to partial commutations, usually under some restrictions on the set I. We first consider the case where D is a clique and then the more general case where D is transitive and finally the case where I is transitive.

5.1 The case where D is a clique

We first consider the case where D consists of a clique (a complete subgraph) and some isolated vertices. To simplify, we shall ignore the isolated vertices and just say in this case that D is a clique. An example is represented below, with $A = \{a, b, c, d, e\}$.



In this case, it is not too hard to modify the proofs of Theorems 4.2 and 4.4 to obtain the following results:

Theorem 5.1 *Let I be a partial commutation such that D is a clique. If L is a group language, then $[L]_I$ is a polynomial of group languages.*

Proof. Let L be a group language, let $\pi : A^* \to G$ be its syntactic morphism and let $R = \pi^{-1}(1)$. We also denote by B the set of vertices of the clique D and by C the set $A \setminus B$. For instance, in our example, we get $B = \{a, b, c\}$ and $C = \{d, e\}$. We claim that the language $K = [L]_I$ satisfies $K = [K]_{\to R}$. Let $u \in K$ and let $r \in R$. Let us write u as $u_0b_1u_1 \cdots b_ku_k$, where $b_1, \ldots, b_k \in B$ and $u_0, \ldots, u_k \in C^*$. If $u = xy$, there is an index i and a factorisation $u_i = u_i'u_i''$ such that $x = u_0b_1u_1 \cdots b_iu_i'$ and $y = u_i''b_{i+1}u_{i+1} \cdots b_ku_k$.

Since $u \in K$, there exists a word $v \in L$ such that $u \sim_I v$. It follows that $\pi_B(u) \sim_I \pi_B(v)$ and since the restriction of I to $B \times B$ is the equality, one can write v as $v_0b_1v_1 \cdots b_kv_k$ with $v_0, \ldots, v_k \in C^*$. Further, since $\pi_C(u) \sim_I \pi_C(v)$ and since the restriction of I to $C \times C$ is a total commutation, one has $u_0u_1 \cdots u_k \sim_I v_0v_1 \cdots v_k$.

Consider the word $w = (v_0b_1v_1 \cdots v_{i-1}b_i)r(v_ib_{i+1} \cdots b_kv_k)$. Since $\pi(r) = 1$, one gets

$$\begin{aligned}\pi(w) &= \pi(v_0b_1v_1 \cdots v_{i-1}b_i)\pi(r)\pi(v_ib_{i+1} \cdots b_kv_k)\\ &= \pi(v_0b_1v_1 \cdots v_{i-1}b_i)\pi(v_ib_{i+1} \cdots b_kv_k) = \pi(v)\end{aligned}$$

and hence $w \in L$. Further, since the letters of C commute with any other letter and since $u_0u_1 \cdots u_k \sim_I v_0v_1 \cdots v_k$, one gets

$$\begin{aligned}w &\sim_I b_1 \cdots b_irb_{i+1} \cdots b_kv_0 \cdots v_k\\ &\sim_I b_1 \cdots b_irb_{i+1} \cdots b_ku_0 \cdots u_k \sim_I xry\end{aligned}$$

It follows that $w \sim_I xry$ and hence $xry \in K$, which proves the claim. The result now follows from Corollary 3.12. □

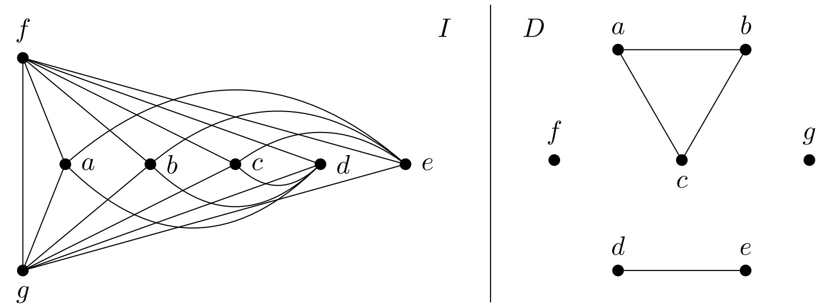
Corollary 5.2 *Let I be a partial commutation such that D is a clique. If L is a polynomial of group languages, then $[L]_I$ is a polynomial of group languages.*

Proof. The proof is similar to that of Theorem 4.4. □

5.2 The case where D is transitive

In this section we extend the results of Section 5.1 to the more general case where D is transitive. This is equivalent to requiring that $A^*/{\sim_I}$ is isomorphic to a direct product of free monoids.

For instance, if $A = \{a, b, c, d, e, f, g\}$, and I and D are the relations represented below, one has $A^*/{\sim_I} = \{a, b, c\}^* \times \{d, e\}^* \times \{f\}^* \times \{g\}^*$.



The proof we present is totally different from that of Theorem 5.1, which does not seem to generalize easily to the transitive case. Suppose that $A^*/{\sim_I} = A_1^* \times \cdots \times A_k^*$. Denote by π_j the projection from A^* onto A_j^*, which is the morphism defined by

$$\pi_j(a) = \begin{cases} a & \text{if } a \in A_j \\ 1 & \text{otherwise} \end{cases}$$

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 $\pi_I(u) = \pi_I(v)$. In particular, recall that $[L]_I$ is regular if and only if $\pi_I(L)$ is a recognisable subset of $A_1^* \times \cdots \times A_k^*$.

Proposition 5.3 *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

$$[L]_I = \bigcup_{1 \leqslant i \leqslant n} L_{i,1} \sqcup\!\sqcup \cdots \sqcup\!\sqcup L_{i,k} \tag{6}$$

Proof. Let K be the right hand side of (6). We first show that $[L]_I$ is a subset of K. Let $u \in [L]_I$. Then there is a word $v \in L$ such that $u \sim_I v$. Let, for $1 \leqslant j \leqslant k$, $v_j = \pi_j(v)$. Then $v \in v_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup v_k$ and thus $(v_1, \ldots, v_k) \in \pi_I(L)$. Therefore, one has $(v_1, \ldots, v_k) \in L_{i,1} \times \cdots \times L_{i,k}$ for some $i \in \{1, \ldots, n\}$. Now, since $u \sim_I v$, the projections of u and v on each A_j^* coincide. It follows that $u \in v_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup v_k$ and hence $u \in L_{i,1} \sqcup\!\sqcup \cdots \sqcup\!\sqcup L_{i,k}$ and finally $u \in K$.

To prove the opposite inclusion, consider a word $u \in K$. Then one has $u \in L_{i,1} \sqcup\!\sqcup \cdots \sqcup\!\sqcup L_{i,k}$ for some $i \in \{1, \ldots, n\}$. Therefore, there exist some words $v_1 \in L_{i,1}$, $\ldots$, $v_k \in L_{i,k}$ such that $u \in v_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup v_k$. Now, since $(v_1, \ldots, v_k) \in L_{i,1} \times \cdots \times L_{i,k}$, one gets by (5) $(v_1, \ldots, v_k) \in \pi_I(L)$. Consequently, there exists a word $v \in L$ such that $\pi_I(v) = (v_1, \ldots, v_k)$, that is $v \in v_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup v_k$. It follows that the projections of u and v on each A_j^* coincide and hence $u \sim_I v$. Thus $u \in [L]_I$. □

We adapt an argument from [6, Proposition 9.6] to compute $\pi_I(L)$ in the special case of a group language. Let $\pi : A^* \to G$ be the syntactic morphism of a group language L.

Proposition 5.4 *Let $N = k|G|^{k+2}$ and, for $1 \leqslant i \leqslant k$, let $R_i = A_i^* \cap \pi^{-1}(1)$. Then the following formula holds:*

$$\pi_I(L) = \bigcup (u_1 \uparrow R_1) \times \cdots \times (u_k \uparrow R_k) \tag{7}$$

where the union runs over the set E of k-tuples of words $(u_1, \ldots, u_k) \in \pi_I(L)$ such that $|u_1|, \ldots, |u_k| \leqslant N$.

Proof. First observe that the conditions

$$(u_1 \ldots, u_k) \in \pi_I(L) \quad \text{and} \quad L \cap (u_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup u_k) \neq \emptyset$$

are equivalent. We shall use freely this remark in the remainder of the proof.

Let K be the right member of (7). We first prove that K is a subset of $\pi_I(L)$. If t is a k-tuple of K, there is a k-tuple $(u_1 \ldots, u_k) \in E$ such that

$$t = (r_{1,0}a_{1,1}r_{1,1} \cdots a_{1,n_1}r_{1,n_1}, \ldots, r_{k,0}a_{k,1}r_{k,1} \cdots a_{k,n_k}r_{k,n_k})$$

where, for $1 \leqslant i \leqslant k$, $u_i = a_{i,1} \cdots a_{i,n_i}$ and $r_{i,j} \in R_i$ for $0 \leqslant j \leqslant n_i$.

Since $(u_1, \ldots, u_k) \in E$, there exists a word $u \in L$ such that $\pi_I(u) = (u_1, \ldots, u_k)$.

Thus u belongs to $u_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup u_k$. Let us replace each letter $a_{i,j}$ in u by the word $r_{i,j-1}a_{i,j}$ if $j < n_i$ and by $r_{i,n_i-1}a_{i,n_i}r_{i,n_i}$ if $j = n_i$. Let us do this operation for $1 \leqslant i \leqslant k$ and $1 \leqslant j \leqslant n_i$. Since $\pi(r_{i,j}) = 1$ for all i, j, the resulting word v has the following properties:

(1) for $1 \leqslant i \leqslant k$, $\pi_i(v) = r_{i,0}a_{i,1}r_{i,1} \cdots a_{i,n_i}r_{i,n_i}$ and hence $\pi_I(v) = t$,

(2) $\pi(v) = \pi(u)$ and thus $v \in L$.

It follows that $t \in \pi_I(L)$ and therefore K is a subset of $\pi_I(L)$.

In the opposite direction, consider a k-tuple $t = (u_1, \ldots, u_k) \in \pi_I(L)$. We prove that $t \in K$ by induction on $|t| = |u_1| + \ldots + |u_k|$. First assume that $|t| \leqslant N$. Then $t \in E$ and thus $t \in (u_1 \uparrow R_1) \times \cdots \times (u_k \uparrow R_k)$, since $1 \in R_i$ for $1 \leqslant i \leqslant k$. It follows that t belongs to K.

We may now assume that $|t| > N$. By assumption, there is a word $u \in L$ such that $\pi_I(u) = (u_1, \ldots, u_k)$. First suppose that, for some i, u contains a factor of length $\geqslant |G|$ in A_i^*. Then by Lemma 3.8, this factor contains a nonempty factor in R_i and thus $u = u'xu''$ with $x \in R_i \cap A^+$. It follows by Lemma 3.5 that $u'u'' \in L$. Further, x is also a factor of u_i, so that $u_i = u_i'xu_i''$. Let $t' = \pi_I(u'u'')$. Then $t' = (u_1, \ldots, u_{i-1}, u_i'u_i'', u_{i+1}, \ldots, u_k)$ and since $|t'| < |t|$, one gets $t' \in K$ by the induction hypothesis. Therefore, there is a k-tuple $(v_1, \ldots, v_k) \in E$ such that $t' \in (v_1 \uparrow R_1) \times \cdots \times (v_k \uparrow R_k)$. In particular, $u_i'u_i'' \in v_i \uparrow R_i$ and by Lemma 3.6, $u_i = u_i'xu_i'' \in v_i \uparrow R_i$. It follows that $t \in (v_1 \uparrow R_1) \times \cdots \times (v_k \uparrow R_k)$ and hence $t \in K$.

Suppose now that u has no factor of length $\geqslant |G|$ in A_i^*. Let us factorize u as

$$u = u_{1,1}u_{1,2} \cdots u_{1,k}u_{2,1} \cdots u_{2,k} \cdots u_{n,1} \cdots u_{n,k}$$

where, for $1 \leqslant j \leqslant n$ and $1 \leqslant i \leqslant k$, $u_{j,i} \in A_i^*$ and $u_{j,1} \cdots u_{j,k} \neq 1$. For instance, if $A_1 = \{a, b\}$, $A_2 = \{c\}$ and $A_3 = \{d, e\}$, the factorization of the word $cabddabcade$ would be $(1)(c)(1)(ab)(1)(dd)(ab)(c)(1)(a)(1)(de)$. Since u has no factor of length $\geqslant |G|$ in A_i^*, the length of each word $u_{i,j}$ is strictly less than $|G|$. On the other hand, $|u| = |t| > N$ and thus $n > |G|^{k+1}$. Note that

$$\pi_I(u) = (u_{1,1} \cdots u_{n,1}, u_{1,2} \cdots u_{n,2}, \ \ldots \ , u_{1,k} \cdots u_{n,k})$$

Let, for $1 \leqslant r \leqslant n$, g_r be the element of the group G^{k+1} defined by

$$g_r = (\pi(u_{r,1}), \pi(u_{r,2}), \ldots, \pi(u_{r,k}), \pi(u_{r,1}u_{r,2} \cdots u_{r,k}))$$

By Lemma 3.4, applied to the group G^{k+1}, there exist two indices i and j, with $i \leqslant j \leqslant |G|^{k+1}$ such that $g_i \cdots g_j = (1, \ldots, 1)$ which means that for $1 \leqslant s \leqslant k$, $u_{i,s} \cdots u_{j,s} \in R_s$ and that $(u_{i,1}u_{i,2} \cdots u_{i,k}) \cdots (u_{j,1}u_{j,2} \cdots u_{j,k}) \in \pi^{-1}(1)$. Now, since $u \in L$, it follows by Lemma 3.5 that

$$(u_{1,1} \cdots u_{1,k}) \cdots (u_{i-1,1} \cdots u_{i-1,k})(u_{j+1,1} \cdots u_{j+1,k}) \cdots (u_{n,1} \cdots u_{n,k}) \in L$$

Therefore the k-tuple

$$(u_{1,1} \cdots u_{i-1,1}u_{j+1,1} \cdots u_{n,1}, \ \ldots \ , u_{1,k} \cdots u_{i-1,k}u_{j+1,k} \cdots u_{n,k})$$

belongs to $\pi_I(L)$ and by the induction hypothesis, also belongs to K. It follows by Lemma 3.6 that $(u_{1,1} \cdots u_{n,1}, u_{1,2} \cdots u_{n,2}, \ \ldots \ , u_{1,k} \cdots u_{n,k})$ belongs to K. Therefore $\pi_I(L) = K$. $\square$

Theorem 5.5 *Let I be a partial commutation such that D is transitive. If L is a group language, then $[L]_I$ is a polynomial of group languages.*

Proof. It follows from Proposition 5.4 that if L is a group language, then $\pi_I(L) = \bigcup_{1 \leqslant i \leqslant n} L_{i,1} \times \cdots \times L_{i,k}$, where each language $L_{i,j}$ is a polynomial of group languages. Since $\mathrm{Pol}(\mathcal{G})$ is closed under shuffle [7], the result now follows directly from (6). □

Theorem 5.6 *Let I be a partial commutation such that D is transitive. If L is a polynomial of group languages, then $[L]_I$ is also a polynomial of group languages.*

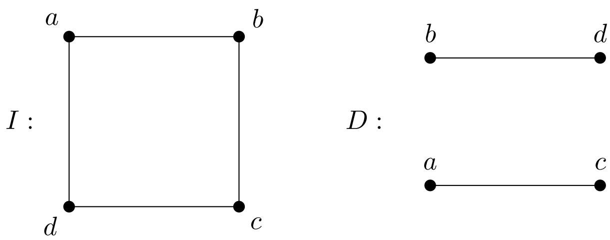
Proof. Since $\mathrm{Pol}(\mathcal{G})$ is closed under shuffle, it suffices, by Proposition 5.3, to prove that if $L \in \mathrm{Pol}(\mathcal{G})$, then $\pi_I(L)$ is a finite union of languages of the form $L_1 \times \cdots \times L_k$, where $L_i \in \mathrm{Pol}(\mathcal{G})(A_i^*)$ for $1 \leqslant i \leqslant k$.

Since π_I is a morphism, it preserves union and product. Therefore it suffices to prove the result if L is of the form $L_0 a_1 L_1 \cdots a_n L_n$, where $L_0, \ldots, L_n$ are group languages. Theorem 5.5 shows that the result holds for the languages $L_0, L_1, \ldots, L_n$, since they are group languages. Further, if a is a letter, then $\pi_I(a) = (1, \ldots, 1, a, 1, \ldots, 1)$, where the i-th component is a if and only if $a \in A_i$. It follows that $\pi_I(L_0 a_1 L_1 \cdots a_n L_n)$ is a finite union of languages of the form $R_1 \times \cdots \times R_k$, where each language R_i is a product of the form $S_0 c_1 S_1 \cdots c_r S_r$, with $S_0, \ldots, S_r \in \mathrm{Pol}(\mathcal{G})(A_i^*)$ and each c_j is either a letter of A_i or the empty word. But since $\mathrm{Pol}(\mathcal{G})$ is closed under product and marked product, R_i belongs to $\mathrm{Pol}(\mathcal{G})(A_i^*)$. □

This result cannot be extended to the languages of $\mathcal{W}$. Indeed we exhibit a partial commutation such that D is transitive and a language of $\mathcal{W}$ such that $[L]_I$ is not regular.

Example 5.1 Consider the alphabet $A = \{a, b, c, d\}$ and the partial commutation relation I (with D transitive) defined by

$$ab \sim_I ba \qquad ad \sim_I da \qquad bc \sim_I cb \qquad cd \sim_I dc$$



Consider the language

$$L = (abcd)^* + A^*aaA^* + A^*bbA^* + A^*ccA^* + A^*ddA^* + A^*ababA^* + A^*bcbcA^* + A^*cdcdA^* + A^*dadaA^*$$

We first show that L belongs to $\mathcal{W}$ and next that $[L]_I$ is not regular.

Let $(M, \leqslant)$ be the syntactic ordered monoid of L. A short computation, using the software `Semigroupe 2.0` shows that M is an aperiodic monoid with zero, containing 170 elements grouped into 4 regular $\mathcal{J}$-classes and some nonregular $\mathcal{J}$-classes. These regular $\mathcal{J}$-classes comprise the singleton $\{1\}$, the minimal ideal $\{0\}$, a unique 0-minimal $\mathcal{J}$-class with 12 $\mathcal{R}$-classes and 12 $\mathcal{L}$-classes and the regular $\mathcal{J}$-class D represented below:

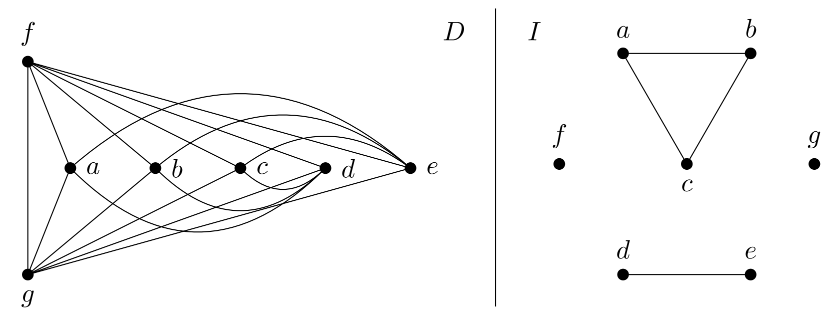
* $bcda$	$bcdab$	bc	bcd
cda	* $cdab$	$cdabc$	cd
da	dab	* $dabc$	$dabcd$
$abcda$	ab	abc	* $abcd$

The presentation of M computed by `Semigroupe` has 116 relations and cannot be reproduced here. Similarly, we shall not give the syntactic order in detail, but we mention that the relation $0 \leqslant x$ holds for all $x \in M$. It follows that if x and y are mutually inverse elements of M such that 0 belongs to the submonoid generated by x and y, then $(xy0xy)^\omega = 0$ and Condition $(*)$ defining $\mathbf{W}$ is trivially satisfied. This covers the trivial case $x = y = 1$ and the cases where x and y belong to the minimal ideal or to the unique 0-minimal ideal. The only remaining case occurs when both x and y belong to D. If x and y are both equal to the same idempotent e of D, Condition $(*)$ is also trivially satisfied. The remaining possibilities for the pair (x, y) are $(abcda, bcd)$, $(bcdab, cda)$, (ab, cd), $(abc, dabcd)$, (bc, da) and $(cdabd, dab)$. But in all these cases, one gets either $x^2 = 0$ or $y^2 = 0$ and again, Condition $(*)$ is trivially satisfied.

We now show that the language $[L]_I$ is not regular by showing that its syntactic congruence has infinite index. For each $n \geqslant 0$, set $x_n = (ac)^n$. We claim that if $i \neq j$, then $x_i \not\sim_{[L]_I} x_j$. Indeed, setting $z_i = (bd)^i$, we get $x_i z_i = (ac)^i(bd)^i \in [L]_I$ since $(abcd)^i \in L$ and $(abcd)^i \sim_I (ac)^i(bd)^i$, but $x_j z_i = (ac)^j(bd)^i \notin [L]_I$ since no word u in L satisfies $(ac)^j(bd)^i \sim_I u$. This proves the claim.

5.3 The case where I is transitive

We now consider the case where I is transitive, which amounts to requiring that $A^*/{\sim_I}$ is isomorphic to a free product of free commutative monoids. For instance, if $A = \{a, b, c, d, e, f, g\}$, and I and D are the relations represented below, $A^*/{\sim_I}$ is isomorphic to the free product $\mathbb{N}^3 * \mathbb{N}^2 * \mathbb{N} * \mathbb{N}$.



Theorem 5.7 *Let L be a language of $\mathcal{W}(A^*)$ and let I be a transitive partial commutation. Then $[L]_I$ is a regular language.*

Proof. Let $\mathcal{P} = \{A_1, \ldots, A_k\}$ be the partition of A such that $A^*/{\sim_I}$ is isomorphic to the free product $\mathbb{N}^{A_1} * \cdots * \mathbb{N}^{A_k}$. For instance, on our example, this partition would be $\bigl\{\{a, b, c\}, \{d, e\}, \{f\}, \{g\}\bigr\}$.

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,

$$\text{for all } u \in A^*,\ q\cdot u \in F \text{ implies } p\cdot u \in F.$$

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), let us set

$$K_{p,q} = \{u \in A^* \mid p\cdot u \leqslant q\}$$

We now create a transition from p to q labelled by

$$R_{p,q} = \bigcup_{1\leqslant i\leqslant k} \bigl[K_{p,q} \cap A_i^*\bigr]$$

Let x be a word such that $q_0\cdot x = p$.

Lemma 5.8 *The following formula holds:*

$$K_{p,q} = \bigcap_{q\cdot y\in F} x^{-1}Ly^{-1} \tag{8}$$

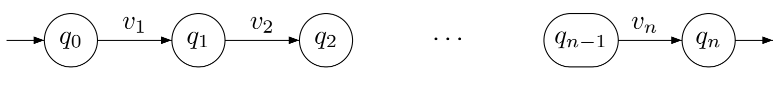
Proof. If $u \in K_{p,q}$, then $p\cdot u \leqslant q$ and thus $q_0\cdot xu \leqslant q$. Therefore, if $q\cdot y \in F$, then $q_0\cdot xuy \in F$ by the definition of $\leqslant$, whence $xuy \in L$ and $u \in x^{-1}Ly^{-1}$.

In the opposite direction, suppose that $u \in x^{-1}Ly^{-1}$ for all words y such that $q\cdot y \in F$. Let us show that $p\cdot u \leqslant q$. Indeed, if $q\cdot y \in F$, then $u \in x^{-1}Ly^{-1}$, whence $xuy \in L$ and $(p\cdot u)\cdot y \in F$. Since this holds for any y such that $q\cdot y \in F$, we have $p\cdot u \leqslant q$ and hence $u \in K_{p,q}$. □

It is well known that a regular language has only finitely many quotients of the form $x^{-1}Ly^{-1}$. Since $\mathcal{W}$ is closed under quotients, it follows from (8) that $K_{p,q}$ belongs to $\mathcal{W}(A^*)$. Since $\mathcal{W}$ is closed under total commutation by Theorem 4.6, $R_{p,q}$ is also in $\mathcal{W}(A^*)$.

We claim that $\mathcal{B}$ recognises $[L]_I$. Let $u \in [L]_I$. Let us factorise u as $u = u_1 \cdots u_n$ where all the letters of each u_i belong to the same class of $\mathcal{P}$, but the letters of two consecutive u_i belong to different classes of $\mathcal{P}$. Continuing our example, the factorisation of $acbadebcbagdefg$ would be $(acba)(de)(bcba)(g)(de)(f)(g)$. Since $u \in [L]_I$, there exist some words $v_1, \ldots, v_n$ such that $u_1 \sim v_1, \ldots, u_n \sim v_n$ and $v_1 \cdots v_n \in L$.

Let $q_1 = q_0\cdot v_1$, $q_2 = q_1\cdot v_2$, $\ldots$, $q_n = q_{n-1}\cdot v_n$. Since $v_1 \cdots v_n$ belongs to L, q_n is a final state.



Now, it follows from the definition of the sets $R_{p,q}$ that $u_1 \in R_{q_0,q_1}$, ..., $u_n \in R_{q_{n-1},q_n}$. Consequently u is accepted by $\mathcal{B}$.

In the opposite direction, consider a word u accepted by $\mathcal{B}$ and let

be a successful path of $\mathcal{B}$ labelled by u, such that the letters of each u_i belong to a single class of the partition $\mathcal{P}$. Thus q_n is a final state and according to the definition of $\mathcal{B}$, there exist some words $v_1, \ldots, v_n$ such that $u_1 \sim v_1$, ..., $u_n \sim v_n$ and $q_0 \cdot v_1 \leqslant q_1$, $q_1 \cdot v_2 \leqslant q_2$, ..., $q_{n-1} \cdot v_n \leqslant q_n$. Setting $v = v_1 \cdots v_n$, it follows that $q_0 \cdot v \leqslant q_n$. Now, by the definition of the order $\leqslant$, the condition $q_n \in F$ implies $q_0 \cdot v \in F$ and hence $v \in L$. It follows that $u \sim_I v$ and thus $u \in [L]_I$. Thus $[L]_I$ is regular. □

We do not know whether $[L]_I$ also belongs to $\mathcal{W}(A^*)$. However, the proof of Theorem 5.7 can be adapted to prove another result. Let (A_1, I_1), ..., (A_k, I_k) be the connected components of the graph (A, I) and put, for $1 \leqslant j \leqslant k$,

$$D_j = \{(a,b) \in A_j \times A_j \mid (a,b) \notin I_j\}$$

Theorem 5.9 *Suppose that, for $1 \leqslant j \leqslant k$, (A_j, D_j) is transitive. Then, if L is a polynomial of group languages, $[L]_I$ is regular.*

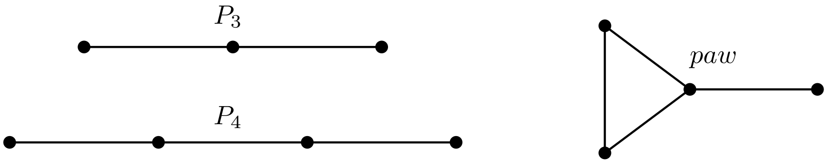
Proof. The proof is similar to that of Theorem 5.7 and we just indicate the modifications needed to extend the result to this more general case.

Let $\mathcal{P} = \{A_1, \ldots, A_k\}$ be the partition of A such that $A^*/{\sim_I}$ 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

$$R_{p,q} = \bigcup_{1 \leqslant j \leqslant k} \left[K_{p,q} \cap A_j^*\right]_{I_j}$$

Formula (8) shows that if $L \in \mathrm{Pol}(\mathcal{G})(A^*)$, then the language $K_{p,q}$ is also in $\mathrm{Pol}(\mathcal{G})(A^*)$. Since $\mathrm{Pol}(\mathcal{G})$ is a positive variety of languages, it is closed under inverse of morphisms. In particular, if $\imath$ denotes the identity map from A_j^* into A^*, one has $K_{p,q} \cap A_j^* = \imath^{-1}(K_{p,q})$ and thus $K_{p,q} \cap A_j^*$ belongs to $\mathrm{Pol}(\mathcal{G})(A_j^*)$. If (A_j, D_j) is transitive, it follows from Theorem 5.6 that $R_{p,q}$ is in $\mathrm{Pol}(\mathcal{G})(A^*)$. The rest of the proof is in unchanged. □

There is a simple graph theoretic interpretation of the condition on I given in the statement of Theorem 5.9. We adopt a `standard graph terminology` and denote respectively by P_3, P_4 and *paw* the graphs represented below:



The graph co-P_3 is the complement of the graph P_3.

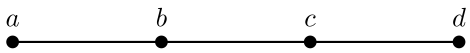
Let us recall a few definitions from graph theory. The *distance* between two vertices of a graph is the number of edges in a shortest path connecting them. The *diameter* of a graph is the greatest distance between two vertices of the graph. Let G and H be two graphs. Let us say that a graph G is H*-free* if there is no subgraph of G isomorphic to H.

Proposition 5.10 *Let I be a partial commutation, let (A_1, I_1), ..., (A_k, I_k) be the connected components of the graph (A, I) and let (A_j, D_j) be the complement graph of (A_j, I_j). Then the following conditions are equivalent:*

(1) *for $1 \leqslant j \leqslant k$, (A_j, D_j) is transitive,*

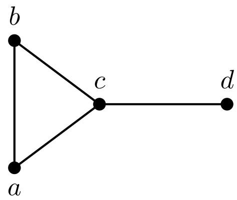
(2) *the graph (A, I) is (P_4, paw)-free.*

Proof. (1) implies (2). Suppose that (1) is satisfied but (2) is not. If there is a subgraph of (A, I) isomorphic to P_4, say



Then the four vertices a, b, c, d are in the same connected component, say (A_j, I_j). However, (a, d) and (d, b) are in (A_j, D_j) but (a, b) is not. This contradicts the fact that (A_j, D_j) is transitive.

Suppose now there is a subgraph of (A, I) isomorphic to *paw*, say



Again, (a, d) and (d, b) are in (A_j, D_j), but (a, b) is not. This contradicts the fact that (A_j, D_j) is transitive.

(2) implies (1). First observe that (A_j, D_j) is transitive if and only if the graph (A_j, I_j) is (co-P_3)-free. Suppose that (A, I) is (P_4, paw)-free. Then every graph (A_j, I_j) is a connected (P_4, paw)-free graph. It is proved in [29] that any connected paw-free graph is either triangle-free or (co-P_3)-free. Therefore it suffices to show that if a connected graph G is (triangle, P_4)-free, then it is co-P_3-free. We first prove a lemma.

Lemma 5.11 *Let G be a (triangle, P_4)-free graph. Then the diameter of G is $\leqslant 2$.*

Proof. Suppose that the diameter of G is $\geqslant 3$ and let a and d be two vertices at distance 3. Then there are two vertices b and c and edges (a, b), (b, c) and (c, d). Since G is triangle-free, there is no edge (a, c) nor (b, d). But since G is P_4-free, there is necessarily an edge (a, d), a contradiction. □

Suppose now that G contains a copy of co-P_3: an edge (a, b), a vertex c such that nor (c, a) nor (c, b) are edges of G. Since G is connected, there is path from c to a. Since the diameter of G is $\leqslant 2$ the length of this path is exactly 2, say $(c, d), (d, a)$. Now, since G is triangle-free, (d, b) is not an edge and $(c, d), (d, a), (a, b)$ form a subgraph isomorphic to P_4, a contradiction. $\square$

We can now state the last result of this section.

Theorem 5.12 *Let L be a polynomial of group languages. If the graph (A, I) is (P_4, paw)-free, then $[L]_I$ is regular.*

One may wonder whether under the conditions of Theorem 5.12, the language $[L]_I$ is a polynomial of group languages. The following example gives a negative answer to this question.

Example 5.2 Let $A = \{a, b, c\}$ and let I be the partial commutation defined by $ab \sim_I ba$. Let L be the set of words having an even number of subwords equal to ab. Then L is a group language. We claim that $[L]_I$ is not a polynomial of group languages. Indeed, one has $aab \in L$, whence $aba \in [L]_I$. However, for each $n > 0$, one has $abc^n a \notin [L]_I$. It follows by [32, Theorem 7.1] that $[L]_I$ is not a polynomial of group languages.

Example 5.2 also shows that $\mathrm{Pol}(\mathcal{G})$ is not closed under partial commutations.

6 Closure under semi-commutation

In this section, we study some properties of semi-commutations. We start by giving some combinatorial properties of semi-commutations.

6.1 Combinatorial properties of semi-commutations

Let A be an alphabet and let I be a semi-commutation on A. Given a subset B of A, we denote by π_B the projection from A^* onto B^*, which is the morphism defined by $\pi_B(a) = a$ if $a \in B$ and $\pi_B(a) = 1$ otherwise. The following result can be found in [11, Lemma 12.3.8].

Proposition 6.1 *Let $u, v \in A^*$. Then $u \xrightarrow{*}_I v$ if and only if, for each subset B of A such that $|B| \leqslant 2$, $\pi_B(u) \xrightarrow{*}_I \pi_B(v)$.*

Proposition 6.1 reduces the study of derivations to the case of a two-letter alphabet $A = \{a, b\}$. Let I be a semi-commutation on A and suppose that $u \xrightarrow{*}_I v$. If I is empty, then of course $u = v$. If I is the full commutation, then u and v are commutatively equivalent. The remaining case $I = \{(a, b)\}$ was first studied by Clerbout, Latteux and Roos [11, Lemma 12.7.2]. Let us denote by $p_n(u)$ the prefix of length n of a word u.

Lemma 6.2 *Let $A = \{a, b\}$ and let $I = \{(a, b)\}$. One has $u \xrightarrow{*}_I v$ if and only if $|v|_a = |u|_a$, $|v|_b = |u|_b$ and for each nonnegative integer n, $|p_n(u)|_b \leqslant |p_n(v)|_b$.*

We are now back to the general case. Let a be a letter. We need a few elementary lemmas about derivations of the form $w \xrightarrow{*}_I xa^k y$, where w, x, y are words.

Lemma 6.3 *Let x, y, u, v be words. If $uav \xrightarrow{*}_I xay$, then $ua^kv \xrightarrow{*}_I xa^ky$ for all $k > 0$.*

Proof. It suffices to observe that if $ab \to_I ba$ [$ba \to_I ab$], then $a^kb \xrightarrow{*}_I ba^k$ [$ba^k \xrightarrow{*}_I a^kb$]. Now, given a derivation from uav to xay, one can obtain a derivation from ua^kv to xa^ky by substituting for each rule $ab \to_I ba$ [$ba \to_I ab$] used in the original derivation the rule $a^kb \xrightarrow{*}_I ba^k$ [$ba^k \xrightarrow{*}_I a^kb$]. □

The next lemma is the counterpart for semi-commutations of [12, Corollary 1.2].

Lemma 6.4 *Let $u_0, u_1, v_0, v_1 \in A^*$ with $|u_0|_a = |v_0|_a$. If $u_0au_1 \xrightarrow{*}_I v_0av_1$, then $u_0u_1 \xrightarrow{*}_I v_0v_1$.*

Proof. Suppose that the conditions of the lemma are satisfied. Let $u = u_0au_1$ and $v = v_0av_1$. By Proposition 6.1 it suffices to prove that

$$\pi_B(u_0u_1) \xrightarrow{*}_I \pi_B(v_0v_1) \tag{9}$$

for each subset B of A such that $|B| \leqslant 2$. The result is trivial if $|B| \leqslant 1$. Suppose that $|B| = 2$. Projecting the relation $u \xrightarrow{*}_I v$ onto B yields

$$\begin{cases} \pi_B(u_0)a\pi_B(u_1) \xrightarrow{*}_I \pi_B(v_0)a\pi_B(v_1) & \text{if } a \in B \\ \pi_B(u_0)\pi_B(u_1) \xrightarrow{*}_I \pi_B(v_0)\pi_B(v_1) & \text{if } a \notin B \end{cases} \tag{10}$$

This gives immediately (9) if $a \notin B$ and thus we may assume that $B = \{a, b\}$. Further, since we only work with the projection onto B of the words u_0, u_1, v_0 and v_1, it just remains to prove our lemma for the case $A = B = \{a, b\}$.

There are two easy cases. If $(a, b) \notin I$ and $(b, a) \notin I$, then $u_0au_1 = v_0av_1$, and since $|u_0|_a = |v_0|_a$, one gets $u_0 = v_0$ and $u_1 = v_1$, which proves the lemma. If a and b commute, then $u_0au_1 \sim_I v_0av_1$, and thus $u_0u_1 \xrightarrow{*}_I v_0v_1$.

The remaining case is $(a, b) \in I$ and $(b, a) \notin I$ — the case $(a, b) \notin I$ and $(b, a) \in I$ is symmetric since $|u_0|_a = |v_0|_a$ implies $|u_1|_a = |v_1|_a$. Applying Lemma 6.2 to the derivation $u_0au_1 \xrightarrow{*}_I v_0av_1$ with $n = |u_0|$, we get $|u_0|_b \leqslant |p_n(v)|_b$ and thus $|p_n(v)|_a \leqslant |u_0|_a = |v_0|_a$. It follows that $n \leqslant |v_0|$, that is, $|u_0| \leqslant |v_0|$.

In order to prove $u_0u_1 \xrightarrow{*}_I v_0v_1$, it suffices now to show that $|p_n(u_0u_1)|_b \leqslant |p_n(v_0v_1)|_b$ for each n. If $n \leqslant |u_0|$, then $|p_n(u_0u_1)|_b = |p_n(u)|_b \leqslant |p_n(v)|_b = |p_n(v_0v_1)|_b$ since $|u_0| \leqslant |v_0|$. If $n \geqslant |v_0|$, then $|p_n(u_0u_1)|_b = |p_{n+1}(u)|_b \leqslant |p_{n+1}(v)|_b = |p_n(v_0v_1))|_b$.

Finally, suppose that $|u_0| < n < |v_0|$. This choice of n gives $|p_{n+1}(u)|_a > |p_{n+1}(v)|_a$ and thus $|p_{n+1}(u)|_b \leqslant |p_{n+1}(v)|_b - 1$. It follows that $|p_n(u_0u_1)|_b = |p_{n+1}(u)|_b \leqslant |p_{n+1}(v)|_b - 1 = |p_n(v_0v_1))|_b$, which completes the proof. □

The following result of Clerbout and Latteux [10, Lemma 3] can be now derived from Lemma 6.4.

Corollary 6.5 *If $uav \xrightarrow{*}_I wa$ with $|v|_a = 0$, then $uv \xrightarrow{*}_I w$ and one has $(a, b) \in I$ for each letter b occurring in v.*

6.2 A small class closed under semi-commutation

We have seen in Theorem 2.4 that the classes Pol($\mathcal{J}$) and Pol($\mathcal{C}om$) are closed under any semi-commutation. We now exhibit another small class enjoying this property. Recall that a language belongs to Pol($\mathcal{I}$) if and only if it is a *shuffle ideal*, that is, a language of the form $L \sqcup\!\sqcup A^*$ for some L. Let $\mathcal{J}^-$ be the class of all languages closed under taking subwords. Note that a language of A^* belongs to $\mathcal{J}^-$ if and only if its complement belongs to Pol($\mathcal{I}$). The dual version of [31, Theorem 6.4] provides a syntactic characterization of $\mathcal{J}^-$.

Proposition 6.6 *A language belongs to $\mathcal{J}^-$ if and only if* 1 *is the minimum of its ordered syntactic monoid.*

The class $\mathcal{J}^-$ is a positive variety of languages. The corresponding variety of ordered monoids, denoted by $\mathbf{J}^-$, is defined by the identity $1 \leqslant x$.

Theorem 6.7 *The class $\mathcal{J}^-$ is closed under any semi-commutation.*

Proof. Let L be a language of A^* closed under taking subwords and let I be a semi-commutation on A. Let $u \in L$. We claim that if $u \xrightarrow{*}_I v$, then for each subword v' of v, there is a subword u' of u such that $u' \xrightarrow{*}_I v'$. We prove the claim by induction on the length n of the derivation from u to v. If $n = 0$, then $u = v$ and the result is trivial. Suppose that $n > 0$. Then there is a word w which derives from u in $n-1$ steps and such that $w \rightarrow_I v$. Then $w = xaby$ and $v = xbay$ for some $(a, b) \in I$. Let v' be a subword of v. If v' is a subword of xay or of xby, then it is also a subword of w and we can conclude by induction. Let us now assume that $v' = x'bay'$ for some subword x' of x and some subword y' of y. Let $w' = x'aby'$. Then w' is a subword of w and by the induction hypothesis, there is a subword u' of u such that $u' \xrightarrow{*}_I w'$. Since $w' \rightarrow_I v'$, we get $u' \xrightarrow{*}_I v'$, which proves the claim. It follows that $[L]_I$ is closed under taking subwords, which proves the theorem. □

6.3 Semi-commutations and product

Let us start with a useful consequence of Theorem 2.3.

Corollary 6.8 *Let I be a semi-commutation on A and let $\mathcal{L}$ be a set of regular languages on A^*. If, for each language L of $\mathcal{L}$, $[L]_I$ is regular, then for each language L of Pol($\mathcal{L}$), $[L]_I$ is regular.*

Proof. Suppose that, for each language L of $\mathcal{L}$, $[L]_I$ is regular. We claim that for each language L of Pol($\mathcal{L}$), $[L]_I$ is regular. Since, for each family $(L_j)_{j \in J}$ of languages, one has

$$\Big[\bigcup_{j \in J} L_j\Big]_I = \bigcup_{j \in J} [L_j]_I \tag{11}$$

it suffices to establish the result for a language L of the form $L_0 a_1 L_1 \cdots a_n L_n$, where $L_0, \ldots, L_n \in \mathcal{L}$ and $a_1, \ldots, a_n$ are letters. Now, since $[a]_I = \{a\}$ for each letter a, the result follows directly from Theorem 2.3. □

This result should be compared with the following result on group languages.

Theorem 6.9 *Let I be semi-commutation on A. If, for each group language K of A^*, $[K]_I$ is a polynomial of group languages, then for each polynomial of group languages L of A^*, $[L]_I$ is a polynomial of group languages.*

The short proof below was communicated to us by Pierre-Cyrille Héam.

Proof. Suppose that for each group language K of A^*, $[K]_I$ is a polynomial of group languages. Let now L be a polynomial of group languages of A^*. By Corollary 6.8, $[L]_I$ is regular. Further, [32, Theorem 7.1] shows that L is open in the pro-group topology, which means that L is a (possibly infinite) union of group languages. By assumption, if K is a group languages, then $[K]_I$ is a polynomial of group languages and hence is open. It follows that $[L]_I$ is a union of open sets and hence is also open. Therefore $[L]_I$ is an open regular language and by [32, Theorem 7.1] again, it is a polynomial of group languages. □

Note that Theorem 6.9 gives another proof of Theorem 5.6.

Theorem 2.3 shows that if $[L_1]_I, \ldots, [L_n]_I$ are regular languages, then the language $[L_1 \cdots L_n]_I$ is regular. A careful analysis of the proof given in [10], which has been often rediscovered [2, 3, 9, 19, 20, 21], leads to a more precise result. Let us introduce some convenient notation. Let I be a semi-commutation on A and let $L_1, \ldots, L_n$ be languages of A^*. Let $B = A \times \{1, \ldots, n\}$ and let $\varphi : B^* \to A^*$ be the morphism defined by $\varphi(a, i) = a$ for all $a \in A$. For $1 \leqslant i \leqslant n$, let $A_i = A \times \{i\}$ and let $\gamma_i : B^* \to B^*$ be the morphism defined by $\gamma_i(a, i) = (a, i)$ and $\gamma_i(a, j) = 1$ if $j \neq i$. Let also $\pi_i = \varphi \circ \gamma_i$. Thus π_i is the morphism from B^* to A^* defined by $\pi_i(a, i) = a$ and $\pi_i(a, j) = 1$ if $j \neq i$. Finally, consider the semi-commutation on B defined by

$$J = \Big\{ \big((a, i), (b, j)\big) \mid (a, b) \in I \text{ and } i, j \in \{1, \ldots, n\} \Big\}$$

Observe that if $u, v \in B^*$ and $u \xrightarrow{*}_J v$, then $\varphi(u) \xrightarrow{*}_I \varphi(v)$. We shall need an auxiliary result. Let

Lemma 6.10 *If $v \in [A_1^* \cdots A_n^*]_J$ then $\gamma_1(v) \cdots \gamma_n(v) \xrightarrow{*}_J v$.*

Proof. We prove the result by induction on the length of v. If $|v| = 0$, the result is trivial. Suppose that $|v| > 0$ and that $v \in [A_1^* \cdots A_n^*]_J$. Let us write v as $v'(a, j)$ with $v' \in B^*$ and $(a, j) \in B$. For $1 \leqslant i \leqslant n$, let u_i be a word of A_i^* such that $u_1 \cdots u_n \xrightarrow{*}_J v'(a, j)$. Since (a, j) belongs to A_j, the letter (a, j) must occur in u_j and one can write u_j as $u_j'(a, j)u_j''$ where u_j'' contains no occurrence of (a, j). Now since

$$u_1 \cdots u_{j-1} u_j'(a, j) u_j'' \cdots u_n \xrightarrow{*}_J v'(a, j), \tag{12}$$

one has by Corollary 6.5 $((a, j), (b, k)) \in J$ for each letter (b, k) of $u_j'' u_{j+1} \cdots u_n$. Note that for $j + 1 \leqslant i \leqslant n$, the letters occurring in u_i are exactly the letters occurring in $\gamma_i(v')$. Thus

$$(a, j)\gamma_{j+1}(v') \cdots \gamma_n(v') \xrightarrow{*}_J \gamma_{j+1}(v') \cdots \gamma_n(v')(a, j)$$

Corollary 6.5 applied to (12) shows that $u_1 \cdots u_{j-1}u'_j u''_j \cdots u_n \xrightarrow{*}_J v'$ and thus $v' \in [A_1^* \cdots A_n^*]_J$. By the induction hypothesis, one has $\gamma_1(v') \cdots \gamma_n(v') \xrightarrow{*}_J v'$ and one finally obtains

$$\begin{aligned}\gamma_1(v) \cdots \gamma_n(v) &= \gamma_1(v') \cdots \gamma_{j-1}(v')\gamma_j(v')(a,j)\gamma_{j+1}(v') \cdots \gamma_n(v') \\ &\xrightarrow{*}_J \gamma_1(v') \cdots \gamma_{j-1}(v')\gamma_j(v')\gamma_{j+1}(v') \cdots \gamma_n(v')(a,j) \xrightarrow{*}_J v'(a,j) = v\end{aligned}$$

which concludes the induction step. □

Proposition 6.11 *The following formula holds:*

$$[L_1 \cdots L_n]_I = \varphi\Big([A_1^* \cdots A_n^*]_J \cap \bigcap_{1 \leqslant i \leqslant n} \pi_i^{-1}([L_i]_I)\Big) \tag{13}$$

Proof. Let $R = [A_1^* \cdots A_n^*]_J \cap \bigcap_{1 \leqslant i \leqslant n} \pi_i^{-1}([L_i]_I)$. We claim that if $v \in [L_1 \cdots L_n]_I$ then $v \in \varphi(R)$. First there exist some words $u_1 \in L_1, \ldots, u_n \in L_n$ such that $u_1 \cdots u_n \xrightarrow{*}_I v$. Let $\bar{u}_i$ be the unique word of A_i^* such that $\pi_i(\bar{u}_i) = u_i$. Let $u = u_1 \cdots u_n$ and let $\bar{u} = \bar{u}_1 \cdots \bar{u}_n$. The derivation $u \xrightarrow{*}_I v$ can be coded by a sequence of transpositions on $\{1, \ldots, |u|\}$ indicating which occurrences of the letters of u are permuted. For instance, if $I = \{(a,c),(b,c),(d,c),(d,a)\}$, $u_1 = da$ and $u_2 = bac$, one has the derivation

$$dabac \xrightarrow{(4,5)}_I dabca \xrightarrow{(1,2)}_I adbca \xrightarrow{(3,4)}_I adcba \xrightarrow{(2,3)}_I acdba \xrightarrow{(1,2)}_I cadba$$

Now starting from the derivation $u \xrightarrow{*}_I v$, we build a derivation $\bar{u} \xrightarrow{*}_J \bar{v}$ with exactly the same coding. On our example, we woud get

$$(d,1)(a,1)(b,2)(a,2)(c,2) \xrightarrow{*}_J (c,2)(a,1)(d,1)(b,2)(a,2)$$

By construction, $\varphi(\bar{v}) = v$. From $\bar{u} \xrightarrow{*}_J \bar{v}$, we get $\pi_i(\bar{u}) \xrightarrow{*}_I \pi_i(\bar{v})$, that is, $u_i \xrightarrow{*}_I \pi_i(\bar{v})$ for $1 \leqslant i \leqslant n$. It follows that $\pi_i(\bar{v}) \in [L_i]_I$ and $\bar{v} \in \pi_i^{-1}([L_i]_I)$. Further since $\bar{u} \in A_1^* \cdots A_n^*$ and $\bar{u} \xrightarrow{*}_J \bar{v}$, we also get $\bar{v} \in [A_1^* \cdots A_n^*]_J$, which proves the claim.

Suppose now that $v = \varphi(\bar{v})$ for some $\bar{v} \in R$. By Proposition 6.10, one has $\gamma_1(\bar{v}) \cdots \gamma_n(\bar{v}) \xrightarrow{*}_J \bar{v}$. Since $\pi_i = \varphi \circ \gamma_i$ and $v = \varphi(\bar{v})$, it follows that $\pi_1(\bar{v}) \cdots \pi_n(\bar{v}) \xrightarrow{*}_I v$. But since $\bar{v} \in \pi_i^{-1}([L_i]_I)$, one has $\pi_i(\bar{v}) \in [L_i]_I$ and there exists $u_i \in L_i$ such that $u_i \xrightarrow{*}_I \pi_i(\bar{v})$. It follows that $u_1 \cdots u_n \xrightarrow{*}_I \pi_1(\bar{v}) \cdots \pi_n(\bar{v})$ and finally $u_1 \cdots u_n \xrightarrow{*}_I v$. This proves that $v \in [L_1 \cdots L_n]_I$. □

We can now state the following variation on Theorem 2.3.

Corollary 6.12 *If $[L_1]_I, \ldots, [L_n]_I$ are languages of $\mathcal{W}$, then $[L_1 \cdots L_n]_I$ is also in $\mathcal{W}$.*

Proof. The language $A_1^* \cdots A_n^*$ is clearly closed under taking subwords and thus belongs to $\mathcal{J}^-$. By Theorem 6.7, $[A_1^* \cdots A_n^*]_J$ also belongs to $\mathcal{J}^-$ and hence to $\mathcal{W}$, since $\mathcal{J}^-$ is contained in $\mathcal{W}$. Since $\mathcal{W}$ is a positive variety closed under length-preserving morphisms, Proposition 6.11 shows that $[L_1 \cdots L_n]_I$ belongs to $\mathcal{W}$. □

6.4 The two-letter case

When $A = \{a, b\}$, one can rework the proof of Theorem 4.2 to get the following result.

Theorem 6.13 *If L is a group language on the alphabet $\{a, b\}$, then $[L]_I$ is a polynomial of group languages for any semi-commutation I.*

Proof. It suffices to treat the case $I = \{(a, b)\}$ since the case of the full commutation follows from Theorem 4.2. Let $\pi : \{a, b\}^* \to G$ be the syntactic morphism of L, let $R = \pi^{-1}(1)$ and let $K = [L]_I$. We claim that $K = [K]_{\to_R}$. It suffices to prove that if $xy \in K$ and $r \in R$, then $xry \in K$. Since $xy \in K$, there exists a word $v \in L$ such that $v \xrightarrow{*}_I xy$. By Lemma 6.2, one gets

$$|v|_a = |xy|_a \quad \text{and} \quad |v|_b = |xy|_b \tag{14}$$

and for each integer n,

$$|p_n(v)|_b \leqslant |p_n(xy)|_b \tag{15}$$

Let us write v as ps with $|p| = |x|$ and $|s| = |y|$. We claim that $prs \xrightarrow{*}_I xry$. Indeed, (14) gives immediately $|prs|_a = |xry|_a$ and $|prs|_b = |xry|_b$. Further, if $n \leqslant |p|$, one has by (15), $|p_n(prs)|_b = |p_n(ps)|_b \leqslant |p_n(xy)|_b = |p_n(xry)|_b$. In particular for $n = |x|$ one gets $|p|_b \leqslant |x|_b$. Therefore, if $|p| \leqslant n \leqslant |pr|$, one has

$$|p_n(prs)|_b = |p_n(pr)|_b = |p|_b + |p_{n-|x|}(r)|_b \leqslant \\ |x|_b + |p_{n-|x|}(r)|_b = |p_n(xr)|_b = |p_n(xry)|_b$$

and in particular for $n = |pr|$, $|pr|_b \leqslant |xr|_b$. Finally, if $|pr| \leqslant n$, one obtains

$$|p_n(prs)|_b = |p_{n-|r|}(ps)|_b + |r|_b \leqslant |p_{n-|r|}(xy)|_b + |r|_b = |p_n(xry)|_b$$

which proves the claim by Lemma 6.2.

Now since $\pi(r) = 1$, one gets

$$\pi(prs) = \pi(p)\pi(r)\pi(s) = \pi(ps)$$

Therefore $prs \in L$ and $xry \in K$, as required. It follows by Corollary 3.12 that K is a polynomial of group languages. □

A direct application of Theorem 6.9 now gives:

Corollary 6.14 *If L is a polynomial of group languages on the alphabet $\{a, b\}$, then $[L]_I$ is a polynomial of group languages for any semi-commutation I.*

Example 5.2 shows that Theorem 6.13 does not extend to a three letter alphabet.

6.5 A partial result

We still do not know whether the closure of a group language L under a semi-commutation I is a regular language. The partial result proved in this section shows that a potential counterexample might be difficult to find. Indeed we show that the generators of the syntactic monoid of $[L]_I$ always have finite order.

Let A be an alphabet and let I be a semi-commutation on A. Let a be a letter of A, let a be a new letter and let $A = A \cup \{a\}$. We define a new semi-commutation I by setting

$$I = I \cup \{(a, b) \mid (a, b) \in I\}$$

We also denote by $\gamma : A^* \to A^*$ the morphism defined by $\gamma(a) = a$ and $\gamma(b) = b$ for all $b \in A$.

Lemma 6.15 *Let $u, v \in A^*$. If $u \xrightarrow{*}_I v$, then $\gamma(u) \xrightarrow{*}_I \gamma(v)$.*

Proof. It suffices to prove that if $u \to_I v$, then $\gamma(u) \to_I \gamma(v)$, but this is obvious. □

Lemma 6.16 *If $u \xrightarrow{*}_I xa^k y$, then $u_0 a u_1 a \cdots u_{k-1} a u_k \xrightarrow{*}_I xa^k y$ for some words $u_0, \ldots, u_k$ such that $u = u_0 a u_1 a \cdots u_{k-1} a u_k$.*

Proof. It suffices to underline the word a^k in $xa^k y$ and to trace back the underlined letters in the derivation from u to $xa^k y$. □

We now prove the following result.

Theorem 6.17 *Let L be a group language of A^* and let n be the size of its syntactic monoid. For each semi-commutation I on A, there exists an integer k such that, for any letter $a \in A$, $a^k \sim_{[L]_I} a^{k+n}$.*

Proof. Let $\pi : A^* \to G$ be the syntactic morphism of L and let $n = |G|$. Let N be the integer given by Corollary 3.3 applied to n. We claim that for any letter $a \in A$, $a^{N+1-n} \sim_{[L]_I} a^{N+1}$. Let $g = \pi(a)$.

Suppose that $xa^{N+1-n}y \in [L]_I$. Then there exists a word w of L such that $w \xrightarrow{*}_I xa^{N+1-n}y$. Since w contains at least one occurrence of a, one can write $w = uav$ for some words u and v. Now Lemma 6.3 shows that $ua^{n+1}v \xrightarrow{*}_I xa^{N+1}y$. Since G is a finite group, one has $g^n = 1$ by Lagrange's theorem, whence $\pi(uav) = \pi(ua^{n+1}v)$. Thus the words uav and $ua^{n+1}v$ have the same syntactic image by π and hence $ua^{n+1}v \in L$. Therefore $xa^{N+1}y \in [L]_I$.

Conversely, assume that $xa^{N+1}y \in [L]_I$ and let u be a word of L such that $u \xrightarrow{*}_I xa^{N+1}y$. By Lemma 6.16, there is a factorization $u = u_0 a u_1 a \cdots u_N a u_{N+1}$ such that $u_0 a u_1 a \cdots u_N a u_{N+1} \xrightarrow{*}_I xa^{N+1}y$. Just like in the proof of Theorem 4.1, by applying Corollary 3.3 to the sequence of words $u_0 a$, $u_1 a$, $\ldots$, $u_N a$, one can find a sequence $0 \leqslant i_0 < i_1 < \cdots < i_n \leqslant N$ such that $\pi(u_{i_0} a \cdots a u_{i_1 - 1}) = \pi(u_{i_1} a \cdots a u_{i_2 - 1}) = \ldots = \pi(u_{i_{n-1}} a \cdots a u_{i_n - 1}) = g^{-1}$. Let us set $r = u_0 a \cdots u_{i_0 - 1} a$, $s = u_{i_n} a \cdots a u_{N+1}$ and

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

Let also $r = \gamma(r)$, $s = \gamma(s)$ and $v = \gamma(v)$. Following again the proof of Theorem 4.1, one gets $\pi(u) = \pi(v)$ and thus $v \in L$. Now, applying n times Lemma 6.4 in an appropriate way yields $v \xrightarrow{*}_I xa^{N+1-n}y$. It follows by Lemma 6.15 that $v \xrightarrow{*}_I xa^{N+1-n}y$ and thus $xa^{N+1-n}y \in [L]_I$. This proves the claim and the theorem. □

7 Conclusion and open problems

Our results on commutations can be summarized in a nutshell as follows :

(1) Both Pol($\mathcal{G}$) and $\mathcal{W}$ are closed under commutation.

(2) If I transitive and if L is in $\mathcal{W}$, then $[L]_I$ is regular.

(3) If D transitive and if L is a polynomial of group languages, then so is $[L]_I$.

(4) If (A, I) is (P_4, paw)-free and if L is a polynomial of group languages, then $[L]_I$ is regular.

We also proved some partial results on semi-commutations. Many questions remain open.

(1) If L is a group language, is $[L]_I$ always regular? The cases where the graph (A, I) is P_4 or *paw* are especially interesting.

(2) If I is a transitive partial commutation and if L is in $\mathcal{W}$, does $[L]_I$ also belong to $\mathcal{W}$?

(3) If D consists of a single clique and L is in $\mathcal{W}$, is $[L]_I$ regular?

(4) Let $\mathcal{V}$ be smallest variety of languages containing the commutative languages and the group languages. Is Pol($\mathcal{V}$) closed under [partial] commutation?

Acknowledgements

The authors would like to thank Pierre-Cyrille Héam for his elegant proof of Theorem 6.9.

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)*, pp. 399–409, IEEE Pub., Boston (MA), USA, 2001.

[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, G. Guaiana and J.-E. Pin, When does partial commutative closure preserve regularity?, in *ICALP 2008, Part II*, L. Aceto and al. (eds.), pp. 209–220, *Lect. Notes Comp. Sci.* vol. 5126, Springer, Berlin, 2008.

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

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

[8] G. Cécé, P.-C. Héam and Y. Mainier, Clôtures transitives de semi-commutations et model-checking régulier, *Technique et Science Informatiques* **27**,1-2 (2008), 7–28.

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

[10] M. Clerbout and M. Latteux, Semi-Commutations, *Information and Computation* **73** (1987), 59–74.

[11] M. Clerbout, M. Latteux and Y. Roos, Semi-commutations, in *The book of traces*, V. Diekert and G. Rozenberg (eds.), ch. 12, pp. 487–552, World Scientific Publishing Co. Inc., River Edge, NJ, 1995.

[12] R. Cori and D. Perrin, Automates et commutations partielles, *RAIRO Inform. Théor.* **19**,1 (1985), 21–32.

[13] 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.

[14] 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.

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

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

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

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

[19] G. Guaiana, A. Restivo and S. Salemi, Complete subgraphs of bipartite graphs and applications to trace languages, *Theoret. Informatics Appl.* **24** (1990), 409–418.

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

[21] 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.

[22] M. Lothaire, *Combinatorics on words*, Cambridge University Press, Cambridge, 1997. With a foreword by Roger Lyndon and a preface by Dominique Perrin, Corrected reprint of the 1983 original, with a new preface by Perrin.

[23] M. Lothaire, *Algebraic Combinatorics on Words*, *Encyclopedia of Mathematics* vol. 90, Cambridge University Press, 2002.

[24] M. S. L′vov, Commutative closures of regular semigroup languages, *Kibernetika (Kiev)* ,2 (1973), 54–58.

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

[26] 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.

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

[28] 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 (eds.), pp. 337–345, *Lect. Notes Comp. Sci.* vol. 4588, Springer, 2007.

[29] S. Olariu, Paw-free graphs, *Inform. Process. Lett.* **28**,1 (1988), 53–54.

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

[31] J.-E. Pin, A variety theorem without complementation, *Russian Mathematics (Iz. VUZ)* **39** (1995), 80–90.

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

[33] J.-E. Pin and P. Weil, Polynomial closure and unambiguous product, *Theory Comput. Systems* **30** (1997), 383–422. Journal version of the article of ICALP 1995.

[34] J.-E. Pin and P. Weil, Semidirect products of ordered semigroups, *Communications in Algebra* **30** (2002), 149–169.

[35] V. N. Red′ko, On the commutative closure of events, *Dopovidi Akad. Nauk Ukraïn. RSR* **1963** (1963), 1156–1159.

[36] 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.

[37] I. Simon, Piecewise testable events, in *Proc. 2nd GI Conf.*, H. Brackage (ed.), pp. 214–222, *Lecture Notes in Comp. Sci.* vol. 33, Springer Verlag, Berlin, Heidelberg, New York, 1975.