A generic characterization of $\text{Pol}(C)$

Thomas Place$^1$ and Marc Zeitoun$^2$

$^1$Bordeaux University, Labri  
$^2$Bordeaux University, Labri

February 20, 2018

1 Introduction

We investigate the polynomial closure operation ($C \mapsto \text{Pol}(C)$) defined on classes of regular languages. We present an interesting and useful connection relating the separation problem for the class $C$ and the membership problem for its polynomial closure $\text{Pol}(C)$. It was first discovered in [6]. This connection is formulated as an algebraic characterization of $\text{Pol}(C)$ which holds when $C$ is an arbitrary quotienting lattice of regular languages and whose statement is parameterized by $C$-separation. Its main application is an effective reduction from $\text{Pol}(C)$-membership to $C$-separation.

Thus, as soon as one designs a $C$-separation algorithm, this yields “for free” a membership algorithm for the more complex class $\text{Pol}(C)$.

Additionally, we present a second transfer theorem which applies to a smaller class than $\text{Pol}(C)$: the intersection class $\text{Pol}(C) \cap \text{co-Pol}(C)$. This is the class containing all languages $L$ such that both $L$ and its complement belong to $\text{Pol}(C)$. This second transfer theorem is a simple corollary of the first one and was originally formulated in [1]. However it is also stronger: it yields a reduction from $\text{Pol}(C) \cap \text{co-Pol}(C)$-membership to $C$-membership.

2 Preliminary definitions

In this section, we fix the terminology and introduce several objects that we shall need to formulate and prove the results presented in the paper.

2.1 Words and languages

For the whole paper, we fix an arbitrary finite alphabet $A$. We denote by $A^*$ the set of all finite words over $A$, and by $\varepsilon \in A^*$ the empty word. Given two words $u, v \in A^*$, we write $u \cdot v$ (or simply $uv$) their concatenation. A language (over $A$) is a subset of $A^*$. Abusing terminology, we denote by $u$ the singleton language $\{u\}$. It is standard to extend the concatenation operation to languages: given $K, L \subseteq A^*$, we write $KL$ for the language $KL = \{uv \mid u \in K \text{ and } v \in L\}$. Moreover, we also consider marked concatenation, which is less standard. Given $K, L \subseteq A^*$, a marked concatenation of $K$ with $L$ is a language of the form $KaL$ for some $a \in A$.

A class of languages $C$ is simply a set of languages. We say that $C$ is a lattice when $\emptyset \in C$, $A^* \in C$ and $C$ is closed under union and intersection: for any $K, L \in C$, we have $K \cup L \in C$ and $K \cap L \in C$. Moreover, a Boolean algebra is a lattice $C$ which is additionally closed under complement: for any $L \in C$, we have $A^* \setminus L \in C$. Finally, a class $C$ is quotienting if it is closed under quotients. That is, for any $L \in C$ and any word $u \in A^*$, the following properties hold:

$$u^{-1}L \overset{\text{def}}{=} \{w \in A^* \mid uw \in L\} \quad \text{and} \quad Lu^{-1} \overset{\text{def}}{=} \{w \in A^* \mid wu \in L\} \quad \text{both belong to } C.$$
All classes that we consider are quotienting Boolean algebras of regular languages. These are the languages that can be equivalently defined by nondeterministic finite automata, finite monoids or monadic second-order logic. In the paper, we work with the definition by monoids, which we recall now.

**Recognition by a monoid.** A monoid is a set $M$ endowed with an associative multiplication $(s, t) \mapsto s \cdot t$ (we often write $st$ for $s \cdot t$) having a neutral element $1_M$, i.e., such that $1_M s = s = s 1_M$ for every $s \in M$. An idempotent of a monoid $M$ is an element $e \in M$ such that $ee = e$. It is folklore that for any finite monoid $M$, there exists a natural number $\omega(M)$ (denoted by $\omega$ when $M$ is understood) such that for any $s \in M$, the element $s^\omega$ is an idempotent.

We may now explain how to recognize languages with monoids. Observe that $A^*$ is a monoid whose multiplication is concatenation (the neutral element is $\varepsilon$). Thus, we may consider monoid morphisms $\alpha : A^* \to M$ where $M$ is an arbitrary monoid. Given such a morphism and some language $L \subseteq A^*$, we say that $L$ is recognized by $\alpha$ when there exists a set $F \subseteq M$ such that $L = \alpha^{-1}(F)$. It is known that $L$ is regular if and only if it can be recognized by a morphism into a finite monoid.

Moreover, since we consider classes of languages that are not closed under complement (i.e. they are only lattices), we need to work with recognition by ordered monoids. An ordered monoid is a pair $(M, \leq)$ such that for any $s \in M$, they are only lattices, we need to work with recognition by ordered monoids. An ordered monoid is a pair $(M, \leq)$ such that for any $s \in M$, $(M, \leq)$ is compatible with its multiplication: given $s_1 s_2, t_1, t_2 \in M$, if $s_1 \leq t_1$ and $s_2 \leq t_2$, then $s_1 s_2 \leq t_1 t_2$. Furthermore, we say that a subset $F \subseteq M$ is a upper set for $\leq$ when given any $s \in F$ and any $t \in M$ such that $s \leq t$, we have $t \in F$ as well. Consider a morphism $\alpha : A^* \to M$ and $\leq$ an order on $M$ such that $(M, \leq)$ is an ordered monoid. We say that some language $L \subseteq A^*$ is $\leq$-recognized by $\alpha$ when there exists a upper set $F \subseteq M$ for $\leq$ such that $L = \alpha^{-1}(F)$.

**Remark 1.** The key idea behind the definition is that the set of languages which are recognized by $\alpha : A^* \to M$ is necessarily closed under complement: if $L = \alpha^{-1}(F)$, then $A^* \setminus L = \alpha^{-1}(M \setminus F)$. However, this is not the case for the set of languages which are $\leq$-recognized by $\alpha$: while $F$ is an upper set, this need not be the case for $M \setminus F$.

Finally, given any regular language $L$, one may define (and compute) a canonical morphism into a finite monoid which recognizes $L$: the syntactic morphism of $L$. Let us briefly recall its definition. One may associate to $L$ an equivalence $\equiv_L$ over $A^*$: the syntactic congruence of $L$. Given $u, v \in A^*$, $u \equiv_L v$ if and only if $xuy \in L \iff xyv \in L$ for any $x, y \in A^*$. It is known and simple to verify that $\equiv_L$ is a congruence on $A^*$. Thus, the set of equivalence classes $M_L = A^*/\equiv_L$ is a monoid and the map $\alpha_L : A^* \to M_L$ which maps any word to its equivalence class is a morphism. The monoid $M_L$ is called the syntactic monoid of $L$ and $\alpha_L$ its syntactic morphism. Finally, we may define a canonical order relation $\leq_L$ (called syntactic order) on the syntactic monoid $M_L$. Given $s, t \in M_L$, we write $s \leq_L t$ when for any $x, y \in M_L$, $xsy \in \alpha_L(L) \Rightarrow xty \in \alpha_L(L)$. It is simple to verify that $(M_L, \leq_L)$ is an ordered monoid and that $L$ is $\leq_L$-recognized by $\alpha_L$.

It is known that $L$ is regular if and only if $M_L$ is finite (i.e., $\equiv_L$ has finite index): this is Myhill-Nerode theorem. In that case, one may compute the syntactic morphism $\alpha_L : A^* \to M_L$ (and the syntactic order on $M_L$) from any representation of $L$ (such as a finite automaton).

**Membership and separation.** In the paper, we are interested in two decision problems which we define now. Both are parameterized by some class of languages $C$. Given a class of languages $C$, the $C$-membership problem is as follows:

**INPUT:** A regular language $L$.

**OUTPUT:** Does $L$ belong to $C$?

Separation is slightly more involved. Given three languages $K, L_1, L_2$, we say that $K$ separates $L_1$ from $L_2$ if $L_1 \subseteq L$ and $L_2 \cap K = \emptyset$. Given a class of languages $C$, we say that $L_1$ is $C$-separable from $L_2$ if some language in $C$ separates $L_1$ from $L_2$. Observe that when $C$ is not closed under complement (which is the case for all classes investigated in the paper), the definition is not symmetrical: $L_1$ could be $C$-separable from $L_2$ while $L_2$ is not $C$-separable from $L_1$. The separation problem associated to a given class $C$ is as follows:
Theorem 2

Assume first that \( \alpha \leq c \) by definition of \( \leq c \).

Proof. \( \alpha \leq c \) implies \( \alpha \leq c \) for all \( c \) in \( C \).

We finish the section with useful tools that we use to manipulate classes that are finite lattices (i.e. one that contains finitely many languages). Consider a finite lattice \( A \) of all languages: given a fixed class \( C \), obtaining a \( C \)-separation algorithm usually requires a solid understanding of \( C \).

2.2 Factorization forest theorem of Simon

When proving our main theorem, we shall need the factorization forest theorem of Simon which is a combinatorial result about finite monoids. We briefly recall it here. We refer the reader to \([3, 4]\) for more details and a proof.

Consider a finite monoid \( M \) and a morphism \( \alpha : A^* \rightarrow M \). An \( \alpha \)-factorization forest is an ordered unranked tree whose nodes are labeled by words in \( A^* \). For any inner node \( x \) with label \( w \in A^* \), if \( w_1, \ldots, w_n \in A^* \) are the labels of its children listed from left to right, then \( w = w_1 \cdots w_n \). Moreover, all nodes \( x \) in the forest must be of the three following kinds:

- **Leaves** which are labeled by either a single letter or the empty word.
- **Binary inner nodes** which have exactly two children.
- **Idempotent inner nodes** which may have an arbitrary number of children. However, the labels \( w_1, \ldots, w_n \) of these children must satisfy \( \alpha(w_1) = \cdots = \alpha(w_n) = e \) where \( e \) is an idempotent element of \( M \).

Note that an idempotent node with exactly two children is also a binary node. This is harmless.

Given a word \( w \in A^* \), an \( \alpha \)-factorization forest for \( w \) is an \( \alpha \)-factorization forest whose root is labeled by \( w \). The height of a factorization forest is the largest \( h \in \mathbb{N} \) such that it contains a branch with \( h \) inner nodes (a single leaf has height 0). We turn to the factorization forest theorem of Simon: there exists a bound depending only on \( M \) such that any word admits an \( \alpha \)-factorization forest of height at most this bound.

**Theorem 2** \(([3, 5])\). Consider a morphism \( \alpha : A^* \rightarrow M \). For all words \( w \in A^* \), there exists an \( \alpha \)-factorization forest for \( w \) of height at most \( 3|M| - 1 \).

2.3 Finite lattices

We finish the section with useful tools that we use to manipulate classes that are finite lattices (i.e. one that contains finitely many languages). Consider a finite lattice \( C \). One may associate a canonical preorder relation over \( A^* \) to \( C \). The definition is as follows. Given \( w, w' \in A^* \), write \( w \leq_c w' \) if and only if the following holds:

For all \( L \in C \), \( w \in L \Rightarrow w' \in L \).

It is immediate from the definition that \( \leq_c \) is transitive and reflexive, making it a preorder. The relation \( \leq_c \) has many applications. We start with an important lemma, which relies on the fact that \( C \) is finite. We say that a language \( L \subseteq A^* \) is an upper set (for \( \leq_c \)) when for any two words \( u, v \in A^* \), if \( u \in L \) and \( u \leq_c v \), then \( v \in L \).

**Lemma 3.** Let \( C \) be a finite lattice. Then, for any \( L \subseteq A^* \), we have \( L \in C \) if and only if \( L \) is an upper set for \( \leq_c \). In particular, \( \leq_c \) has finitely many upper sets.

**Proof.** Assume first that \( L \in C \). Then, for all \( w \in L \) and all \( w' \) such that \( w \leq_c w' \), we have \( w' \in L \) by definition of \( \leq_c \). Hence, \( L \) is an upper set. Assume now that \( L \) is an upper set. For any word \( w \), we write \( \uparrow w \) for the upper set \( \uparrow w = \{ u \mid w \leq_c u \} \). By definition of \( \leq_c \) \( \uparrow w \) is the intersection of all \( L \in C \) such that \( w \in L \). Therefore, \( \uparrow w \in C \) since \( C \) is a finite lattice (and is therefore closed under intersection). Finally, since \( L \) is an upper set, we have,

\[
L = \bigcup_{w \in L} \uparrow w.
\]
Hence, since $\mathcal{C}$ is closed under union and is finite, $L$ belongs to $\mathcal{C}$. \hfill \Box

We complete this definition with another useful result. When $\mathcal{C}$ is additionally closed under quotients, the canonical preorder $\leq_C$ is compatible with word concatenation.

**Lemma 4.** Let $\mathcal{C}$ be a quotienting lattice. Then, the associated canonical preorder $\leq_C$ is compatible with word concatenation. That is, for any words $u, v, u', v'$,

$$u \leq_C u' \quad \text{and} \quad v \leq_C v' \implies uv \leq_C u'v'.$$

**Proof.** Let $u, u', v, v'$ be four words such that $u \leq_C u'$ and $v \leq_C v'$. We have to prove that $uv \leq_C u'v'$. Let $L \in \mathcal{C}$ and assume that $uv \in L$. We use closure under left quotients to prove that $uv' \in L$ and then closure under right quotients to prove that $u'v' \in L$ which terminates the proof of this direction. Since $uv \in L$, we have $v \in u^{-1} \cdot L$. By closure under left quotients, we have $u^{-1} \cdot L \in \mathcal{C}$, hence, since $v \leq_C v'$, we obtain that $v' \in u^{-1} \cdot L$ and therefore that $uv' \in L$. It now follows that $u \in L \cdot (v')^{-1}$ and that $L \cdot (v')^{-1} \in \mathcal{C}$. Since $u \leq_C u'$, we conclude that $u' \in L \cdot (v')^{-1}$ which means that $u'v' \in L$, as desired. \hfill \Box

## 3 Polynomial closure

In this section, we define the polynomial closure operation defined on classes of languages. It is the main focus of the paper. We also prove a characteristic property of this operation that will be useful in proofs later.

### 3.1 Definition

Given an arbitrary class $\mathcal{C}$, the polynomial closure of $\mathcal{C}$, denoted by $\text{Pol}(\mathcal{C})$, is the smallest class containing $\mathcal{C}$ and closed under marked concatenation and union: for any $H, L \in \text{Pol}(\mathcal{C})$ and $a \in A$, we have $HaL \in \text{Pol}(\mathcal{C})$ and $H \cup L \in \text{Pol}(\mathcal{C})$.

It is not immediate that $\text{Pol}(\mathcal{C})$ has robust closure properties beyond those that are explicitly stated in the definitions. However, it turns out that when $\mathcal{C}$ satisfies robust properties itself, this is the case for $\text{Pol}(\mathcal{C})$ as well. It was shown by Arfi [2] that when $\mathcal{C}$ is a quotienting lattice of regular languages, then $\text{Pol}(\mathcal{C})$ is one as well. Note that this result is not immediate (the difficulty is to prove that $\text{Pol}(\mathcal{C})$ is closed under intersection).

**Theorem 5.** Let $\mathcal{C}$ be a quotienting lattice of regular languages. Then, $\text{Pol}(\mathcal{C})$ is a quotienting lattice of regular languages closed under concatenation and marked concatenation.

We shall obtain an alternate proof of Theorem 5 as a corollary of our main result (i.e. our algebraic characterization of $\text{Pol}(\mathcal{C}s)$).

Finally, we shall consider two additional operations which are defined by building on polynomial closure. Given a class $\mathcal{C}$, we denote by $\text{co-Pol}(\mathcal{C})$ the class containing all complements of languages in $\text{Pol}(\mathcal{C})$: $L \in \text{co-Pol}(\mathcal{C})$ when $A^* \setminus L \in \text{Pol}(\mathcal{C})$. Finally, we also write $\text{Pol}(\mathcal{C}) \cap \text{co-Pol}(\mathcal{C})$ for the class of all languages that belong to both $\text{Pol}(\mathcal{C})$ and $\text{co-Pol}(\mathcal{C})$. The following result is an immediate corollary of Theorem 5.

**Corollary 6.** Let $\mathcal{C}$ be a quotienting lattice of regular languages. Then, $\text{co-Pol}(\mathcal{C})$ is a quotienting lattice of regular languages and $\text{Pol}(\mathcal{C}) \cap \text{co-Pol}(\mathcal{C})$ is a quotienting Boolean algebra of regular languages.

**Proof.** By Theorem 5 $\text{Pol}(\mathcal{C})$ is a quotienting lattice of regular languages. Since quotients commute with Boolean operations, it follows from De Morgan’s laws that $\text{co-Pol}(\mathcal{C})$ is a quotienting lattice of regular languages as well. Consequently, $\text{Pol}(\mathcal{C}) \cap \text{co-Pol}(\mathcal{C})$ is a quotienting lattice of regular languages and since it must be closed under complement by definition, it is actually a quotienting Boolean algebra of regular languages. \hfill \Box
3.2 Characteristic property

We complete the definitions with a property which applies to the polynomial closure of any finite quotienting lattice $C$. Recall that in this case, we associate a canonical preorder $\leq_C$ over $A^*$ (two words are comparable when any language in $C$ containing the first word contains the second word as well). Since $C$ is closed under quotients, $\leq_C$ must be compatible with word concatenation by Lemma 4.

**Proposition 7.** Let $C$ be a finite quotienting lattice. Consider a language $L \subseteq A^*$ in $Pol(C)$. Then, there exist natural numbers $h, p \geq 1$ such that for any $\ell \geq h$ and $u, v, x, y \in A^*$ satisfying $u \leq_C v$, we have,

$$xu^{p\ell+1}y \in L \implies xu^{p\ell}vy^{p\ell}y \in L$$

We now concentrate on proving Proposition 7. We fix the finite quotienting lattice $C$ for the proof. Consider a language $L \subseteq A^*$ in $Pol(C)$. We first need to choose the natural numbers $h, p \geq 1$ depending on $L$ and $C$. We start by choosing $p$ with the following fact.

**Fact 8.** There exists $p \geq 1$ such that for any $m, m' \geq 1$ and $w \in A^*$, $w^pm \leq_C w^pm'$.

**Proof.** Let $\sim$ be the equivalence on $A^*$ generated by $\leq_C$. Since $\leq_C$ is a preorder with finitely many upper sets which is compatible with concatenation (see Lemma 5 and 4), $\sim$ must be a congruence of finite index. Therefore, the set $A^*/\sim$ of $\sim$-classes if a finite monoid. It suffices to choose $p$ as the idempotent power of this finite monoid. \hfill $\square$

It remains to choose $h$. Since $L$ belongs to $Pol(C)$, it is built from languages in $C$ using only union and marked concatenations. It is simple to verify that these two operations commute. Hence, $L$ is a finite union of products having the form:

$$L_0a_1L_1 \cdots a_mL_m,$$

where $a_1, \ldots, a_m \in A$ and $L_0, \ldots, L_m \in C$. We define $n \in \mathbb{N}$ as a natural number such that for any product $L_0a_1L_1 \cdots a_mL_m$ in the union, we have $m \leq n$. Finally, we let,

$$h = 2n + 1$$

It remains to show that $h$ and $p$ satisfy the desired property. Let $\ell \geq h$ and $u, v, x, y \in A^*$ satisfying $u \leq_C v$. We have to show that,

$$xu^{p\ell+1}y \in L \implies xu^{p\ell}vy^{p\ell}y \in L$$

Consequently, we assume that $xu^{p\ell+1}y \in L$. By hypothesis, we know that there exists a product $L_0a_1L_1 \cdots a_mL_m \subseteq L$ with $a_1, \ldots, a_m \in A$, $L_0, \ldots, L_m \in C$ and $m \leq n$ such that $xu^{p\ell+1}y \in L_0a_1L_1 \cdots a_mL_m$. It follows that $xu^{p\ell+1}y$ admits a unique decomposition,

$$xu^{p\ell+1}y = wo_0a_1w_1 \cdots a_my_m$$

such that $w_i \in L_i$ for all $i \leq m$. Recall that by definition $\ell \geq h = 2n + 1 \geq 2m + 1$. Therefore, it is immediate from a pigeon-hole principle argument that an infix $u^p$ of $xu^{p\ell+1}y$ must be contained within one of the infixes $w_i$. In other words, we have the following lemma.

**Lemma 9.** There exists $i \leq m$, $j_1, j_2 < \ell$ such that $j_1 + 1 + j_2 = \ell$ and $x_1, x_2 \in A^*$ satisfying,

- $w_i = x_1u^px_2$.
- $w_0a_1w_1 \cdots a_jx_1 = xu^{pj_1}$.
- $x_2a_{j+1} \cdots a_my_m = u^{pj_2+1}y$. 


We may now finish the proof. By Fact 3, we have the following inequality,

\[ u^p \leq_C u^{p(f+1)} = u^{p(j_1+1+j_2+1)} = u^{p(j_2+1)}u^{p(j_1+1)-1} \]

Moreover, since \( u \leq_C v \) and \( u \leq_C v \) is compatible with concatenation this yields that,

\[ u^p \leq_C u^{p(j_2+1)}v^{p(j_1+1)-1} \]

Using again compatibility with concatenation we obtain,

\[ w_i = x_1u^p x_2 \leq_C x_1u^{p(j_2+1)}v^{p(j_1+1)-1} x_2 \]

Therefore, since \( w_i \in L_i \) which is a language of \( C \), it follows from the definition of \( \leq_C \) that \( x_1u^{p(j_2+1)}v^{p(j_1+1)-1} x_2 \in L_i \). Therefore, since \( w_j \in L_j \) for all \( j \),

\[ w_0a_1w_1 \ldots a_ix_1u^{p(j_2+1)}v^{p(j_1+1)-1}x_2a_{i+1} \ldots a_m w_m \in L_0a_1L_1 \ldots a_mL_m \]

By the last two items in Lemma 9, this exactly says that \( xu^{p(j_2+1)}vu^{p(j_1+1)-1}y \in L_0a_1L_1 \ldots a_mL_m \). Since we have \( L_0a_1L_1 \ldots a_mL_m \subseteq L \) by definition, this implies that \( xu^{p(j_2+1)}vu^{p(j_1+1)-1}y \in L \), finishing the proof.

## 4 Membership for \( Pol(C) \)

In this section, we prove the main theorem of the paper. Given an arbitrary quotienting lattice of regular languages \( C \), \( Pol(C) \)-membership reduces to \( C \)-separation. We state this result in the following theorem.

**Theorem 10.** Let \( C \) be a quotienting lattice of regular languages and assume that \( C \)-separation is decidable. Then \( Pol(C) \)-membership is decidable as well.

**Remark 11.** Theorem 10 is a generalization of a result from [6] which applies only to specific quotienting lattices \( C \) belonging to a hierarchy of classes called the Straubing-Thérien hierarchy. However, let us point out that the main ideas behind the proof are all captured by the special case presented in [6].

This section is devoted to proving Theorem 10. It is based on an algebraic characterization of \( Pol(C) \). This characterization is formulated using equations on the syntactic ordered monoid of the language. These equations are parameterized by a relation on the syntactic monoid: the \( C \)-pairs. As we shall see, computing this relation requires an algorithm for \( C \)-separation which explains the statement of Theorem 10.

We first present the definition of \( C \)-pairs. We then use them to present the algebraic characterization of \( Pol(C) \) and explain why Theorem 10 is an immediate corollary. Finally, we then present a proof of this characterization. It relies on Simon’s factorization forest theorem (Theorem 2).

### 4.1 \( C \)-pairs

Consider a class of languages \( C \), an alphabet \( A \), a finite monoid \( M \) and a surjective morphism \( \alpha : A^* \rightarrow M \). We define a relation on \( M \): the \( C \)-pairs (for \( \alpha \)). Consider a pair \((s, t) \in M \times M \). We say that,

\[ (s, t) \text{ is a } C\text{-pair (for } \alpha \text{) if and only if } \alpha^{-1}(s) \text{ is not } C \text{-separable from } \alpha^{-1}(t) \text{.} \quad (1) \]

**Remark 12.** While we often make this implicit, being a \( C \)-pair depends on the morphism \( \alpha \).

**Remark 13.** While we restrict ourselves to surjective morphisms, observe that the definition makes sense for arbitrary ones. We choose to make this restriction to ensure that we get a reflexive relation, which is not the case when \( \alpha \) is not surjective (if \( s \in M \) has no antecedent \((s, s) \) is not a \( C \)-pair). However this restriction is harmless: we use \( C \)-pairs together with syntactic morphisms which are surjective.
By definition, the set of $\mathcal{C}$-pairs for $\alpha$ is finite: it is a subset of $M \times M$. Moreover, having a $\mathcal{C}$-separation algorithm in hand is clearly enough to compute all $\mathcal{C}$-pairs for any input morphism $\alpha$. While simple, this property is crucial, we state it in the following lemma.

**Lemma 14.** Let $\mathcal{C}$ be a class of languages and assume that $\mathcal{C}$-separation is decidable. Then, given an alphabet $A$, a finite monoid $M$ and a surjective morphism $\alpha : A^* \to M$ as input, one may compute all $\mathcal{C}$-pairs for $\alpha$.

We complete the definition with a few properties of $\mathcal{C}$-pairs. A simple and useful one is that the $\mathcal{C}$-pair relation is reflexive (it is not transitive in general).

**Lemma 15.** Let $\mathcal{C}$ be a class of languages, $A$ an alphabet, $M$ a finite monoid and $\alpha : A^* \to M$ a surjective morphism. Then, the $\mathcal{C}$-pair relation is reflexive: for any $s \in M$, $(s, s)$ is a $\mathcal{C}$-pair.

**Proof.** Given $s \in M$, since $\alpha$ is surjective, we have $\alpha^{-1}(s) \neq \emptyset$. Therefore, $\alpha^{-1}(s) \cap \alpha^{-1}(s) \neq \emptyset$ and we obtain that $\alpha^{-1}(s)$ is not $\mathcal{C}$-separable from $\alpha^{-1}(s)$. This exactly says that $(s, s)$ is a $\mathcal{C}$-pair. 

Finally, we prove that when $\mathcal{C}$ is a quotienting lattice of regular languages (which is the only case that we shall consider), the $\mathcal{C}$-pair relation is compatible with multiplication.

**Lemma 16.** Let $\mathcal{C}$ be a quotienting lattice of regular languages, $A$ an alphabet $M$ a finite monoid and $\alpha : A^* \to M$ a surjective morphism. For any two $\mathcal{C}$-pairs $(s_1, t_1), (s_2, t_2) \in M \times M$, $(s_1, s_2, t_1, t_2)$ is a $\mathcal{C}$-pair as well.

**Proof.** We prove the contrapositive. Assume that $(s_1 s_2, t_1 t_2)$ is not a $\mathcal{C}$-pair. We show that either $(s_1, t_1)$ is not a $\mathcal{C}$-pair or $(s_2, t_2)$ is not a $\mathcal{C}$-pair. By hypothesis, we have a separator $K \in \mathcal{C}$ such that $\alpha^{-1}(s_1 s_2) \subseteq K$ and $K \cap \alpha^{-1}(t_1 t_2) = \emptyset$. We define,

$$H = \bigcap_{w \in \alpha^{-1}(s_2)} Kw^{-1}$$

By definition, $H \in \mathcal{C}$ since $\mathcal{C}$ is a quotienting lattice and contains only regular languages (thus $K$ has finitely many right quotients by the Myhill-Nerode theorem)). Moreover, since $\alpha^{-1}(s_1 s_2) \subseteq K$, one may verify from the definition that $\alpha^{-1}(s_1) \subseteq H$. There are now two cases. If $\alpha^{-1}(t_1) \cap H = \emptyset$ then $H \in \mathcal{C}$ separates $\alpha^{-1}(s_1)$ from $\alpha^{-1}(t_1)$ and we are finished: $(s_1, t_1)$ is not a $\mathcal{C}$-pair. Otherwise, there exists a word $u \in \alpha^{-1}(t_1) \cap H \neq \emptyset$. Let $G = u^{-1}K \in \mathcal{C}$. We claim that $G$ separates $\alpha^{-1}(s_2)$ from $\alpha^{-1}(t_2)$ which concludes the proof: $(s_1, t_1)$ is not a $\mathcal{C}$-pair. Indeed, given $w \in \alpha^{-1}(s_2)$, we have $u \in H \subseteq Kw^{-1}$ which means that $uw \in K$ and therefore that $w \in G = u^{-1}K$. Moreover, assume by contradiction that there exists $v \in \alpha^{-1}(t_2) \cap G$. Since $G = u^{-1}K$, it follows that $uv \in K$. Finally, since $\alpha(u) = t_1$ and $\alpha(v) = t_2$, it follows that $uw \in \alpha^{-1}(t_1 t_2)$. Thus, $uw \in K \cap \alpha^{-1}(t_1 t_2)$ which is a contradiction since this language is empty by hypothesis.

### 4.2 Characterization theorem

We now characterize of $Pol(\mathcal{C})$ when $\mathcal{C}$ is an arbitrary quotienting lattice by a property of the syntactic morphism of the languages in $Pol(\mathcal{C})$. As we announced, the characterization is parametrized by the $\mathcal{C}$-pair relation that we defined above.

**Theorem 17.** Let $\mathcal{C}$ be a quotienting lattice of regular languages and let $L$ be a regular language. Then, the three following properties are equivalent:

1. $L \in Pol(\mathcal{C})$.

2. The syntactic morphism $\alpha_L : A^* \to M_L$ of $L$ satisfies the following property:

$$s^{\omega+1} \leq_L s^\omega ts^\omega \quad \text{for all $\mathcal{C}$-pairs $(s, t) \in M^2_L$.}$$

3. The syntactic morphism $\alpha_L : A^* \to M_L$ of $L$ satisfies the following property:

$$e \leq_L ete \quad \text{for all $\mathcal{C}$-pairs $(e, t) \in M^2_L$ with $e$ idempotent.}$$

(3)
Theorem [17] states a reduction from $Pol(C)$-membership to $C$-separation. Indeed, the syntactic morphism of a regular language can be computed and Equation (2) can be decided as soon as one is able to compute all $C$-pairs (which is equivalent to deciding $C$-separation by Lemma [13]). Hence, we obtain Theorem [10] as an immediate corollary. Moreover, Theorem [5] is also a simple corollary of Theorem [17] (it is straightforward to verify that any class satisfying Item (2) in the theorem has to be a quotienting lattice).

Moreover, observe that one may also use Theorem [17] to obtain a symmetrical characterization for the class $co-Pol(C)$. Recall that $co-Pol(C)$ contains all languages whose complement is in $Pol(C)$. It is straightforward to verify that a language and its complement have the same syntactic monoid but opposite syntactic orders. Therefore, we obtain the following corollary.

Corollary 18. Let $C$ be a quotienting lattice of regular languages and let $L$ be a regular language. Then, the following properties are equivalent:

1. $L \in co-Pol(C)$.

2. The syntactic morphism $\alpha_L : A^* \to M_L$ of $L$ satisfies the following property:
   \[ s^\omega t s^\omega \leq_L s^{\omega+1} \quad \text{for all } C\text{-pairs } (s, t) \in M_L^2. \]  
   (4)

3. The syntactic morphism $\alpha_L : A^* \to M_L$ of $L$ satisfies the following property:
   \[ ete \leq_L e \quad \text{for all } C\text{-pairs } (e, t) \in M_L^2 \text{ with } e \text{ idempotent}. \]  
   (5)

This terminates the presentation of the algebraic characterization of $Pol(C)$. We now turn to its proof.

4.3 Proof of Theorem [17]

We prove Theorem [17]. Let $C$ be a quotienting lattice of regular languages, and let us fix a regular language $L$. Let $\alpha_L : A^* \to M_L$ be its syntactic morphism. We prove that $1) \Rightarrow 2) \Rightarrow 3) \Rightarrow 1)$.

We start with $1) \Rightarrow 2)$: when $L \in Pol(C)$, $\alpha_L$ satisfies Equation (2).

Direction $1) \Rightarrow 2)$

Assume that $L \in Pol(C)$. We have to show that $\alpha_L$ satisfies Equation (2). Given a $C$-pair $(s, t) \in M_L^2$, we have to show that $s^{\omega+1} \leq_L s^\omega t s^\omega$. We first prove the following simple fact.

Fact 19. There exists a finite quotienting lattice $D \subseteq C$ such that $L \in Pol(D)$.

Proof. Since $L \in Pol(C)$, it is built from finitely many languages in $C$ using unions and marked concatenations. We let $F \subseteq C$ as the finite class containing all basic languages in $C$ used in the construction. Moreover, we let $D$ as the smallest quotienting lattice containing $F$. Clearly $D \subseteq C$ since $C$ is a quotienting lattice itself. Moreover, $L \in Pol(D)$ since $D$ contains all languages in $C$ required to build $L$ by definition. It remains to show that $D$ remains finite. By definition, the languages in $D$ are built from those in $F$ by applying unions and intersections. Therefore, since quotients commute with Boolean operations, any language in $D$ is built by applying intersections and unions to languages in $F$. Finally, any regular language has finitely many quotients by Myhill-Nerode theorem. Thus, since $F$ was finite, this is the case for $D$ as well.

We work with the canonical preorder $\leq_D$ over $A^*$ associated to the finite quotienting lattice $D$. Since $(s, t)$ is a $C$-pair, we know that $\alpha^{-1}(s)$ is not $C$-separable from $\alpha^{-1}(t)$. Therefore, since $D \subseteq C$, it follows that $\alpha^{-1}(s)$ is not $D$-separable from $\alpha^{-1}(t)$. Consider the language,

\[ H = \{ v \in A^* \mid u \leq_D v \text{ for some } u \in \alpha^{-1}(s) \} \]

By definition, $H$ is an upper set for $\leq_D$ and therefore belongs to $D$ by Lemma [3]. Moreover, $H$ includes $\alpha^{-1}(s)$ by definition. Consequently, since $\alpha^{-1}(s)$ is not $D$-separable from $\alpha^{-1}(t)$, we
know that \( H \) intersects \( \alpha^{-1}(t) \). This yields \( u \in \alpha^{-1}(s) \) and \( v \in \alpha^{-1}(t) \) such that \( u \leq_D v \). Hence, we may apply Proposition 7 which yields natural numbers \( h, p \geq 1 \) such that for any \( x, y \in A^* \),

\[
x u^{ph \omega + 1} y \in L \quad \Rightarrow \quad x u^{ph \omega} v u^{ph \omega} y \in L
\]

By definition of the syntactic order on \( M_L \), it then follows that,

\[
s^{\omega + 1} = \alpha(u^{ph \omega + 1}) \leq_L \alpha(u^{ph \omega} v u^{ph \omega}) = s^{\omega} t s^{\omega}
\]

This concludes the proof for this direction.

**Direction 2) \( \Rightarrow 3) \)**

Let us assume that the syntactic morphism \( \alpha_L : A^* \to M_L \) of \( L \) satisfies (2). We need to prove that it satisfies (3) as well. Let \((e, t) \in M^2 \) be a \( C \)-pair with \( e \) idempotent. We have to show that \( e \leq_L ete \). Since (2) holds, we know that \( e^{\omega + 1} \leq_L e^{\omega} te^{\omega} \). Moreover, since \( e \) is idempotent, we have \( e = e^{\omega + 1} = e^{\omega} \). Thus, we get \( e \leq_L ete \) as desired.

**Direction 3) \( \Rightarrow 1) \)**

It now remains to prove the harder \( \text{“}(3) \Rightarrow 1)\)” direction of Theorem 17. We use induction to prove that for any finite ordered monoid \((M, \leq)\) and any surjective morphism \( \alpha : A^* \to M \) satisfying (3), any language \( \leq \)-recognized by \( \alpha \) may be constructed from languages of \( C \) using unions and (marked) concatenations (thus showing that it belongs to \( Pol(C) \)). Since \( L \) is \( \leq_L \)-recognized by its syntactic morphism, this ends the proof.

We fix a surjective morphism \( \alpha : A^* \to M \) satisfying (3): for any \( C \)-pair \((e, t) \in M^2 \) with \( e \) idempotent, we have \( e \leq ete \). The proof is based on Simon's factorization forest theorem (see Section 2). We state it in the following proposition.

**Proposition 20.** For all \( h \in \mathbb{N} \) and all \( s \in M \), there exists \( H_{s,h} \in Pol(C) \) such that for all \( w \in A^* \):

- If \( w \in H_{s,h} \) then \( s \leq \alpha(w) \).
- If \( \alpha(w) = s \) and \( w \) admits an \( \alpha \)-factorization forest of height at most \( h \) then \( w \in H_{s,h} \).

Assume for now that Proposition 20 holds. Given \( h = 3|M| - 1 \), for all \( s \in M \), consider the language \( H_{s,h} \in Pol(C) \) associated to \( s \) and \( h \) by Proposition 20. We know from Simon's Factorization Forest theorem (Theorem 2) that all words in \( A^* \) admit an \( \alpha \)-factorization forest of height at most 3\(|M| - 1 \). Therefore, for all \( w \in A^* \) we have,

1. If \( w \in H_{s,h} \) then \( s \leq \alpha(w) \).
2. If \( \alpha(w) = s \) then \( w \in H_{s,h} \).

Let \( L \) be some language \( \leq \)-recognized by \( \alpha \) and let \( F \) be its accepting set. Observe that \( L = \bigcup_{s \in F} H_{s,h} \). Indeed, by Item 2 above, we have \( L \subseteq \bigcup_{s \in F} H_{s,h} \). Moreover, by definition of \( \leq \)-recognizability, \( F \) has to be an upper set, that is, if \( s \in F \) and \( s \leq t \) then \( t \in F \). Hence, Item 1 above implies that \( \bigcup_{s \in F} H_{s,h} \subseteq L \). We conclude that \( L \in Pol(C) \) since it is a union of languages \( H_{s,h} \in Pol(C) \). This finishes the proof of Theorem 17. It now remains to prove Proposition 20.

We begin with a lemma which defines the basic languages in \( C \) that we will use in the construction of our languages in \( Pol(C) \). Note that this is also where we use the fact that (3) holds.

**Lemma 21.** For any idempotent \( e \in M \), there exists a language \( K_e \) belonging to \( C \) (and therefore to \( Pol(C) \)) which satisfies the two following properties,

1. For all \( u \in K_e \), we have \( e \leq e \alpha(u) e \).
2. \( \alpha^{-1}(e) \subseteq K_e \).
Proof. Let $T \subseteq M$ be the set of all elements $t \in M$ such that $(e, t)$ is not a $C$-pair (i.e., $\alpha^{-1}(e)$ is $C$-separable from $\alpha^{-1}(t)$). By definition, for all $t \in T$, there exists a language $G_t \in C$ which separates $\alpha^{-1}(e)$ from $\alpha^{-1}(t)$. We let $K_e = \bigcap_{t \in T} G_t$. Clearly, $K_e \subseteq C$ since $C$ is a quotienting lattice, and is therefore closed under intersection. Moreover, $\alpha^{-1}(e) \subseteq K_e$ since the inclusion holds for all languages $G_t$. Finally, given $u \in K_e$, it is immediate from the definition that $\alpha(u)$ does not belong to $T$ which means that $(e, \alpha(u))$ is a $C$-pair. The first item is now immediate from (3) since $e$ is idempotent.

We may now start the proof of Proposition 20. Let $h \geq 1$ and $s \in M$. We construct $H_{s,h} \in Pol(C)$ by induction on $h$. Assume first that $h = 0$. Note that the nonempty words having an $\alpha$-factorization forest of height at most $0$ are all single letters. We let $B = \{ b \in A \mid \alpha(b) = s \}$. Moreover, we use the language $K_1$ as defined in Lemma 21 for the neutral element $1$ (which is an idempotent). There are two cases depending on whether $s = 1$ or not. If $s \neq 1$, we let,

$$H_{s,0} = \bigcup_{b \in B} K_1 b K_1.$$

Otherwise, when $s = 1$, we let,

$$H_{s,0} = K_1 \bigcup_{b \in B} K_1 b K_1.$$

Note that $H_{s,0} \in Pol(C)$ since we only used marked concatenation and unions and $K_1 \subseteq Pol(C)$ by definition in Lemma 21. We now prove that this definition satisfies the two conditions in Proposition 20. We do the proof for the case when $s \neq 1$ (the other case is similar).

Assume first that $w \in H_{s,0}$, we have to prove that $s \leq \alpha(w)$. By definition $w = ubu'$ with $u, u' \in K_1$ and $b \in B$. Hence, $\alpha(w) = \alpha(u)\alpha(u')$. Since $u, u' \in K_1$, we obtain from the second item in Lemma 21 that $1 \leq \alpha(u)$ and $1 \leq \alpha(u')$. It follows that $s \leq \alpha(u)\alpha(u') = \alpha(w)$.

We turn to the second item. Let $w \in A^*$ such that $\alpha(w) = s$ and $w$ admits an $\alpha$-factorization forest of height at most $0$. Since we assumed that $s \neq 1$, $w$ cannot be empty. We have to prove that $w \in H_{s,0}$. By hypothesis, $w$ is a one letter word $b \in B$. Hence, $w \in K_1 b K_1$ since $e \in K_1$ by the first item in Lemma 21.

Assume now that $h > 0$. There are two cases depending on whether $s$ is idempotent or not. We treat the idempotent case (the other case is essentially a simpler version of the same proof). Hence, we assume that $s$ is an idempotent, that we denote by $e$. We begin by constructing $H_{e,h}$ and then prove that it satisfies the conditions in the proposition. For all $t \in M$, one can use induction to construct $H_{t,h-1} \in Pol(C)$ such that for all $w \in A^*$:

- If $w \in H_{t,h-1}$ then $t \leq \alpha(w)$.
- If $\alpha(w) = t$ and $w$ is empty or admits an $\alpha$-factorization forest of height at most $h - 1$, then $w \in H_{t,h-1}$.

We now define $H_{e,h}$ as the union of three languages. Intuitively, the first one contains the words which are either empty or have an $\alpha$-factorization forest of height at most $h - 1$, the second one, words having an $\alpha$-factorization forest of height $h$ and whose root is a binary node, and the third one, words with an $\alpha$-factorization forest of height $h$ and whose root is an idempotent node.

$$H_{e,h} = H_{e,h-1} \cup \bigcup_{t_1 t_2 = e} \big( H_{t_1,h-1} H_{t_2,h-1} \big) \cup H_{e,h-1} K_e H_{e,h-1}$$

with $K_e$ as defined in Lemma 21.

Note that by definition, $H_{e,h}$ is a union of concatenations of languages in $Pol(C)$ and therefore belongs to $Pol(C)$ itself. We need to prove that it satisfies the conditions of the proposition. Choose some $w \in A^*$ and assume first that $w \in H_{e,h}$. We need to prove that $e \leq \alpha(w)$.

- If $w \in H_{e,h-1}$, then this is by definition of $H_{e,h-1}$.
We first define this new object and then use it to present the characterization of it. It turns out that the former is a simple corollary of the latter: it is obtained via a few algebraic reduction. Intuitively, this second transfer result is much stronger than the previous one. However, unlike the

\[ \alpha \]

\[ \text{Theorem 22.} \]

by Almeida, Bartonová, Klíma and Kunc [1].

\[ \alpha \]

This concludes the proof of Proposition 20.

5 Membership for \( \text{Pol}(C) \cap \text{co-Pol}(C) \)

In this last section, we present a second transfer theorem which applies to the intersection class \( \text{Pol}(C) \cap \text{co-Pol}(C) \). Recall that this denotes the class made of all languages which belong to both \( \text{Pol}(C) \) and \( \text{co-Pol}(C) \).

The membership problem is simpler to handle for \( \text{Pol}(C) \cap \text{co-Pol}(C) \) than it is for \( \text{Pol}(C) \). Recall that using the generic characterization of \( \text{Pol}(C) \) (i.e. Theorem 17) to decide \( \text{Pol}(C) \)-membership requires an algorithm for \( C \)-separation. In other words, we reduced \( \text{Pol}(C) \)-membership to a stronger problem for \( C \): separation. It turns out that deciding membership for \( \text{Pol}(C) \cap \text{co-Pol}(C) \) only requires an algorithm for \( C \)-membership: the same problem is used on both ends of the reduction. Intuitively, this second transfer result is much stronger than the previous one. However, it turns out that the former is a simple corollary of the latter: it is obtained via a few algebraic manipulations on the generic characterization of \( \text{Pol}(C) \) (i.e. Theorem 17). This was first observed by Almeida, Bartonová, Klíma and Kunc [1].

**Theorem 22.** Let \( C \) be a quotienting lattice of regular languages and assume that \( C \)-membership is decidable. Then \( (\text{Pol}(C) \cap \text{co-Pol}(C)) \)-membership is decidable as well.

This section is devoted to proving Theorem 22. Similarly to Theorem 10, the argument is based on an algebraic characterization of \( \text{Pol}(C) \cap \text{co-Pol}(C) \) parametrized by a relation depending on \( C \). However, unlike the \( C \)-pairs that we used in the \( \text{Pol}(C) \)-characterization (i.e. Theorem 17), this new relation can be computed as soon as \( C \)-membership is decidable. We speak of saturated \( C \)-pairs.

We first define this new object and then use it to present the characterization of \( \text{Pol}(C) \cap \text{co-Pol}(C) \).

5.1 Saturated \( C \)-pairs

Consider a class of languages \( C \), an alphabet \( A \), a finite monoid \( M \) and a surjective morphism \( \alpha : A^* \rightarrow M \). We define a new relation on \( M \): the saturated \( C \)-pairs (for \( \alpha \)). Consider a pair \( (s, t) \in M \times M \). We say that,

\[ (s, t) \text{ is a saturated } C \text{-pair (for } \alpha) \]

if and only if

\[ \text{no language } K \in C \text{ recognized by } \alpha \text{ separates } \alpha^{-1}(s) \text{ from } \alpha^{-1}(t) \]
Consider a class $C$ (including those recognized by $\alpha$-morphism. Then, for any $\alpha \in A^*$, the two following properties are equivalent:

1. $\alpha^{-1}(F) \in C$.

2. $F$ is a upper set for the saturated $C$-pair relation: for any $s \in F$ and any $t \in M$ such that $(s,t)$ is a saturated $C$-pair, we have $t \in F$.

Proof. We start with the direction $(1) \Rightarrow (2)$. Assume that $\alpha^{-1}(F) \in C$. Consider $s \in F$ and $t \in M$ such that $(s,t)$ is a saturated $C$-pair, we show that $t \in F$. We proceed by contradiction, assume that $t \notin F$. In that case it is immediate that $\alpha^{-1}(F)$ separates $\alpha^{-1}(s)$ from $\alpha^{-1}(t)$. Since
we have $\alpha^{-1}(F) \in \mathcal{C}$, this contradicts the hypothesis that $(s, t)$ is a saturated $\mathcal{C}$-pair and we are finished.

We turn to the direction $(2) \Rightarrow (1)$. Assume that for any $s \in F$ and any $t \in M$ such that $(s, t)$ is a saturated $\mathcal{C}$-pair, we have $t \in F$. We show that $\alpha^{-1}(F) \in \mathcal{C}$. Consider $s \in F$ and $r \notin F$. By hypothesis, we know that $(s, r)$ is not a saturated $\mathcal{C}$-pair. Thus, we have $G_{s, r} \subseteq M$ such that $\alpha^{-1}(G_{s, r})$ belongs to $\mathcal{C}$ and separates $\alpha^{-1}(s)$ from $\alpha^{-1}(r)$. One may then verify that,

$$\alpha^{-1}(F) = \bigcup_{s \in F} \bigcap_{r \notin F} \alpha^{-1}(G_{s, r})$$

Since $\mathcal{C}$ is a lattice, follows that $\alpha^{-1}(F) \in \mathcal{C}$. This concludes the proof.

We may now further connect the saturated $\mathcal{C}$-pair relation with original $\mathcal{C}$-pair relation. We show that the former is the transitive closure of the latter.

**Lemma 28.** Consider a lattice $\mathcal{C}$, an alphabet $A$, a finite monoid $M$ and a surjective morphism $\alpha : A^* \rightarrow M$. Then, for any $(s, t) \in M \times M$, the following properties are equivalent,

1. $(s, t)$ is a saturated $\mathcal{C}$-pair.
2. There exist $n \in \mathbb{N}$ and $r_0, \ldots, r_{n+1} \in M$ such that $r_0 = s$, $r_{n+1} = t$ and $(r_i, r_{i+1})$ is a $\mathcal{C}$-pair for all $i \leq n$.

**Proof.** We already proved the direction $(2) \Rightarrow (1)$. Indeed, we know from Fact 23 that any $\mathcal{C}$-pair is also a saturated $\mathcal{C}$-pair. Moreover, we showed in Lemma 28 that the saturated $\mathcal{C}$-pair relation is transitive. Therefore, we concentrate on the direction $(1) \Rightarrow (2)$. Let $(s, t)$ be a saturated $\mathcal{C}$-pair. Let $F \subseteq M$ as the smallest subset of $M$ satisfying the two following properties:

1. $s \in F$.
2. For any $\mathcal{C}$-pair $(u, v) \in M \times M$, if $u \in F$, then $v \in F$ as well.

We have $s \in F$ by definition. We show that $\alpha^{-1}(F) \in \mathcal{C}$. By Lemma 27, this will imply that $t \in F$ as well since $(s, t)$ is a saturated $\mathcal{C}$-pair. Thus, $(2)$ holds.

Observe that for any $u \in F$, we may build a language $H_u \in \mathcal{C}$ such that $\alpha^{-1}(u) \subseteq H_u \subseteq \alpha^{-1}(F)$. Indeed, for any $v \notin F$, we know that $(u, v)$ is not a $\mathcal{C}$-pair by definition of $F$. Thus, we have $H_{u, v} \in \mathcal{C}$ which separates $\alpha^{-1}(u)$ from $\alpha^{-1}(v)$. We may now define,

$$H_u = \bigcap_{v \notin F} H_{u, v}$$

Clearly $H_u \in \mathcal{C}$ since $\mathcal{C}$ is a lattice. It now suffices to observe that,

$$\alpha^{-1}(F) = \bigcup_{u \in F} \alpha^{-1}(u) \subseteq \bigcup_{u \in F} H_u \subseteq \alpha^{-1}(F)$$

Thus, $\alpha^{-1}(F) = \bigcup_{u \in F} H_u$ belong to $\mathcal{C}$ since $\mathcal{C}$ is lattice.

Finally, we prove that when $\mathcal{C}$ is a quotienting lattice the saturated $\mathcal{C}$-pair relation is compatible with multiplication.

**Lemma 29.** Let $\mathcal{C}$ be a quotienting lattice of regular languages, $A$ an alphabet $M$ a finite monoid and $\alpha : A^* \rightarrow M$ a surjective morphism. For any two saturated $\mathcal{C}$-pairs $(s_1, t_1), (s_2, t_2) \in M \times M$, $(s_1s_2, t_1t_2)$ is a saturated $\mathcal{C}$-pair as well.

**Proof.** Immediate from Lemma 28 since we already know that the $\mathcal{C}$-pair relation is compatible with multiplication by Lemma 16.


5.2 Characterization theorem

We may now present the announced algebraic characterization of $Pol(C) \cap \co-Pol(C)$ and use it to prove Theorem \ref{thm:characterization}.

**Theorem 30.** Let $C$ be a quotienting lattice of regular languages and $L$ a regular language. Then, the three following properties are equivalent:

1. $L \in Pol(C) \cap \co-Pol(C)$.
2. The syntactic morphism $\alpha_L : A^* \to M_L$ of $L$ satisfies the following property:
   \[ s^{\omega+1} = s^{\omega} t s^{\omega} \quad \text{for all} \quad (s, t) \in M^2 \quad (7) \]
3. The syntactic morphism $\alpha_L : A^* \to M_L$ of $L$ satisfies the following property:
   \[ s^{\omega+1} = s^{\omega} t s^{\omega} \quad \text{for all} \quad (s, t) \in M^2 \quad (8) \]

As announced, Theorem \ref{thm:characterization} states a reduction from $(Pol(C) \cap \co-Pol(C))$-membership to $C$-membership. Indeed, the syntactic morphism of a regular language can be computed and Equation \ref{eq:characterization} can be decided as soon as one is able to compute all saturated $C$-pairs (as we explained, this amounts to deciding $C$-membership). Hence, we obtain Theorem \ref{thm:characterization} as an immediate corollary. We turn to the proof of Theorem \ref{thm:characterization}.

**Proof of Theorem \ref{thm:characterization}** The equivalence $(1) \iff (2)$ follows from Theorem \ref{thm:pol} and Corollary \ref{cor:pol}. Indeed, by definition $L \in Pol(C) \cap \co-Pol(C)$ if and only if $L \in Pol(C)$ and $L \in \co-Pol(C)$. By Theorem \ref{thm:pol} and Corollary \ref{cor:pol} respectively, this is equivalent to $\alpha_L$ satisfying the two following properties:

\[ s^{\omega+1} \leq_L s^{\omega} t s^{\omega} \quad \text{for all} \quad (s, t) \in M^2 \]
\[ s^{\omega+1} \geq_L s^{\omega} t s^{\omega} \quad \text{for all} \quad (s, t) \in M^2 \]

Clearly, when put together, these two equations are equivalent to $(7)$. This concludes the proof of $(1) \iff (2)$.

We now show that $(2) \iff (3)$. The direction $(3) \Rightarrow (2)$ is immediate from Fact \ref{fact:pol}. Indeed, since any $C$-pair is also a saturated $C$-pair, it is immediate that when $(8)$ holds, then $(7)$ holds as well. Therefore, we concentrate on the direction $(2) \Rightarrow (3)$. We assume that $(7)$ holds and prove that this is the case for $(8)$ as well. Consider a saturated $C$-pair $(s, t) \in M^2$. We have to show that $s^{\omega+1} = s^{\omega} t s^{\omega}$.

By Lemma \ref{lem:saturated} we know that there exist $n \in \mathbb{N}$ and $r_0, \ldots, r_{n+1} \in M$ such that $r_0 = s$, $r_{n+1} = t$ and $(r_i, r_{i+1})$ is a $C$-pair for all $i \leq n$. We prove by induction that for all $1 \leq k \leq n+1$, we have,

\[ s^{\omega+1} = s^{\omega} r_k s^{\omega} \]

The case $k = n + 1$ yields the desired result since $r_{n+1} = t$. When $k = 1$, it is immediate from $(7)$ that $s^{\omega+1} = s^{\omega} r_1 s^{\omega}$ since $(s, r_1)$ is a $C$-pair. We now assume that $k > 1$. Using induction, we get that,

\[ s^{\omega+1} = s^{\omega} r_k s^{\omega} \]

Therefore, we obtain,

\[ s^{\omega} = (s^{\omega+1})^{\omega} = (s^{\omega} r_k^{-1} s^{\omega})^{\omega} \]

Since $(r_{k-1}, r_k)$ is a $C$-pair, It is immediate from Lemma \ref{lem:pol} that, $(s^{\omega} r_{k-1}^{-1} s^{\omega}, s^{\omega} r_k s^{\omega})$ is a $C$-pair as well. Thus, it follows from $(7)$ that,

\[ (s^{\omega} r_{k-1}^{-1} s^{\omega})^{\omega+1} = (s^{\omega} r_{k-1}^{-1} s^{\omega})^{\omega} s^{\omega} r_k s^{\omega} (s^{\omega} r_{k-1}^{-1} s^{\omega})^{\omega} \]

Since $s^{\omega+1} = s^{\omega} r_{k-1}^{-1} s^{\omega}$ and $s^{\omega} = (s^{\omega} r_{k-1}^{-1} s^{\omega})^{\omega}$, this yields,

\[ s^{\omega+1} = (s^{\omega+1})^{\omega} = s^{\omega} s^{\omega} r_k s^{\omega} = s^{\omega} r_k s^{\omega} \]

This concludes the proof. \qed
References

[1] Jorge Almeida, Jana Bartonová, Ondrej Klíma, and Michal Kunc. On decidability of intermediate levels of concatenation hierarchies. In Proceedings of the 19th International Conference on Developments in Language Theory, DLT’15, pages 58–70, 2015.

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

[3] Mikolaj Bojańczyk. Factorization forests. In Proceedings of the 13th International Conference on Developments in Language Theory, DLT’09, pages 1–17, Berlin, Heidelberg, 2009. Springer-Verlag.

[4] Thomas Colcombet. Factorization forests for infinite words and applications to countable scattered linear orderings. Theoretical Computer Science, 411(4-5):751–764, 2010.

[5] Manfred Kufleitner. The height of factorization forests. In Proceedings of the 33rd International Symposium on Mathematical Foundations of Computer Science, MFCS’08, pages 443–454, Berlin, Heidelberg, 2008. Springer-Verlag.

[6] Thomas Place and Marc Zeitoun. Going higher in the first-order quantifier alternation hierarchy on words. In Proceedings of the 41st International Colloquium on Automata, Languages, and Programming, ICALP’14, pages 342–353, Berlin, Heidelberg, 2014. Springer-Verlag.

[7] Imre Simon. Factorization forests of finite height. Theoretical Computer Science, 72(1):65–94, 1990.