Separating without any ambiguity

Thomas Place, Marc Zeitoun

Separating without any ambiguity

Thomas Place[1] and Marc Zeitoun[2]

1 **LaBRI Bordeaux University, France**
`tplace@labri.fr`
2 **LaBRI Bordeaux University, France**
`mz@labri.fr`

Abstract

We investigate a standard operator on classes of languages: unambiguous polynomial closure. We show that if $\mathcal{C}$ is a class of regular languages having some mild properties, the membership problem for its unambiguous polynomial closure $UPol(\mathcal{C})$ reduces to the same problem for $\mathcal{C}$. We give a new, self-contained and elementary proof of this result. We also show that unambiguous polynomial closure coincides with alternating left and right deterministic closure. Finally, if additionally $\mathcal{C}$ is finite, we show that the separation and covering problems are decidable for $UPol(\mathcal{C})$.

1998 ACM Subject Classification F.4.3 Formal Languages

Keywords and phrases Regular languages, separation problem, decidable characterizations

Digital Object Identifier 10.4230/LIPIcs.ICALP.2018.

1 Introduction

Most of the interesting classes of regular languages are built using a restricted set of operators. From a class $\mathcal{C}$, one may consider its Boolean closure $Bool(\mathcal{C})$, its polynomial closure $Pol(\mathcal{C})$, and deterministic variants thereof, which yield usually a more elaborate class than $\mathcal{C}$. It is therefore desirable to investigate the operators themselves rather than individual classes.

The *polynomial closure* $Pol(\mathcal{C})$ of a class $\mathcal{C}$ is its closure under union and marked concatenation (a *marked concatenation* of K and L is a language of the form KaL for a letter a). Together with the Boolean closure, it is used to define concatenation hierarchies: starting from a given class (*level 0* in the hierarchy), level $n + \frac{1}{2}$ is the polynomial closure of level n, and level $n + 1$ is the Boolean closure of level $n + \frac{1}{2}$. The importance of these hierarchies stems from the fact that they are the combinatorial counterpart of quantifier alternation hierarchies in logic, which count the number of $\forall/\exists$ alternations needed to define a language.

The main question when investigating a class of languages is whether it is recursive: can we decide whether a given input language belongs to the class? This is the *membership* problem. Despite decades of research on concatenation hierarchies, one knows little about this question. The state of the art is that when level 0 is finite and has some mild properties, membership is decidable for levels $\frac{1}{2}$, 1, $\frac{3}{2}$, and $\frac{5}{2}$ [18, 15, 12, 19]. These results imply those that were obtained previously [3, 2, 22, 10, 11] and even go beyond by investigating the *separation problem*, a generalization of membership. Unlike membership, which takes a single language as input, the separation problem for a class $\mathcal{C}$ takes *two*. It asks whether there exists a third language from $\mathcal{C}$, containing the first and disjoint from the second. Membership is the special case of separation when the input consists of a language and its complement. Although more difficult than membership, separation is also more rewarding. This is witnessed by a transfer theorem [15, 19]: membership for $Pol(\mathcal{C})$ reduces to separation for $\mathcal{C}$. The results on membership quoted above actually come from this theorem and the fact that separation is decidable for $Pol(\mathcal{C})$, $BPol(\mathcal{C})$ and $Pol(BPol(\mathcal{C}))$ when $\mathcal{C}$ is finite with some mild properties.

45th International Colloquium on Automata, Languages, and Programming (ICALP 2018).
Editors: Ioannis Chatzigiannakis, Christos Kaklamanis, Daniel Marx, and Don Sannella;
Article No. ; pp. :1–:35

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Unambiguous closure. Deterministic variants of the polynomial closure are also important. The most classical example is the unambiguous closure, where marked concatenations are required to be unambiguous. A marked concatenation KaL is *unambiguous* if every word w of KaL has a unique factorization $w = w'aw''$ with $w' \in K$ and $w'' \in L$. The *unambiguous closure* $UPol(\mathcal{C})$ of $\mathcal{C}$ is the closure of $\mathcal{C}$ under *disjoint* union and *unambiguous* concatenation. Note that it is not clear on the definition whether $UPol(\mathcal{C})$ is a Boolean algebra, even when $\mathcal{C}$ is.

The class of *unambiguous languages* [21] is the unambiguous polynomial closure of the Boolean algebra generated by languages of the form A^*aA^* (where A is the working alphabet). It is one of the most investigated class. It is robust, with several definitions [23, 4, 6, 5, 24]. Unambiguous polynomial closure also appears in concatenation hierarchies as *intermediate levels*. Pin and Weil [10, 11] have shown that $UPol(\mathcal{C}) = Pol(\mathcal{C}) \cap co\text{-}Pol(\mathcal{C})$, where $co\text{-}Pol(\mathcal{C})$ is the class consisting of all complements of languages in $Pol(\mathcal{C})$.

Contributions. By considering separation, we obtained a better understanding of the results that were already known for this kind of problem and we were able to prove new ones. Aside from the case of unambiguous languages [14], unambiguous polynomial closure was not yet investigated with respect to separation. This is the starting point of this paper: we look for a generic separation result applying to $UPol(\mathcal{C})$, similar to the ones obtained for $Pol(\mathcal{C})$ and $BPol(\mathcal{C})$ in [18]. In this paper, we present such a result: our main theorem states that when $\mathcal{C}$ is finite and satisfies some mild hypotheses (the same as for getting decidability of $Pol(\mathcal{C})$-separation), separation for $UPol(\mathcal{C})$ is decidable. However, as it is usually the case with separation, we also obtain several additional results as a byproduct of our work which improve our understanding of the $UPol$ operator:

- We had to rethink the way membership is classically handled for $UPol(\mathcal{C})$ in order to lift the techniques to separation. This yields a completely new, self-contained and elementary proof that under some natural hypothesis on $\mathcal{C}$, membership for $UPol(\mathcal{C})$ reduces to membership for $\mathcal{C}$. This proof also precisely pinpoints why this result holds for $UPol(\mathcal{C})$ but not $Pol(\mathcal{C})$. More precisely, we show that the languages from $\mathcal{C}$ needed to construct an $UPol(\mathcal{C})$ expression for a language L are all recognized by any recognizer of L.
- We obtain a new proof that $UPol(\mathcal{C})$ is a quotienting Boolean algebra when $\mathcal{C}$ is one.
- We obtain a new proof that $UPol(\mathcal{C}) = Pol(\mathcal{C}) \cap co\text{-}Pol(\mathcal{C})$ using our results on $Pol(\mathcal{C})$ [19].
- We obtain a previously unknown characterization of $UPol(\mathcal{C})$ in terms of alternating left and right deterministic concatenations, which are restricted forms of unambiguous concatenation. A marked concatenation KaL is *left (resp. right) deterministic* when $KaA^* \cap K = \emptyset$ (resp. $A^*aL \cap L = \emptyset$). We prove that $UPol(\mathcal{C})$ coincides with $ADet(\mathcal{C})$, the closure of $\mathcal{C}$ under left and right deterministic concatenation. This was observed in the above particular case, but not known in general.

Related work. $UPol(\mathcal{C})$ was characterized in [8, 9]. However [8] starts from an alternate definition that assumes closure under Boolean operations already. Both papers use elaborate mathematical tools (categories, bilateral kernel, relational morphisms) and both use black boxes (results by Schützenberger [21] in [8] and a very general result of Rhodes [20] in [9]). The reduction from $Pol(\mathcal{C})$-membership to $\mathcal{C}$-membership was also investigated in [1]. While our reduction is direct, the proof if [1] is not, as it uses the nontrivial equality $UPol(\mathcal{C}) = Pol(\mathcal{C}) \cap co\text{-}Pol(\mathcal{C})$.

2 Preliminaries

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

$$u^{-1}L \stackrel{\text{def}}{=} \{w \in A^* \mid uw \in L\} \quad \text{and} \quad Lu^{-1} \stackrel{\text{def}}{=} \{w \in A^* \mid wu \in L\} \quad \text{both belong to } \mathcal{C}.$$

All classes that we consider are quotienting Boolean algebras of regular languages.

Regular languages. These are the languages that can be equivalently defined by non-deterministic finite automata, finite monoids or monadic second-order logic. In the paper, we work with the definition by monoids, which we recall now.

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 $1_M \cdot s = s \cdot 1_M = s$ 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 ω when M is understood) such that for any $s \in M$, the element s^ω is an idempotent.

Our proofs make use of the Green relations [7], which are defined on monoids (we use them as induction parameters). We briefly recall them. Given a monoid M and $s, t \in M$,

$$\begin{array}{ll} s \leqslant_{\mathcal{J}} t & \text{when there exist } x, y \in M \text{ such that } s = xty \\ s \leqslant_{\mathcal{L}} t & \text{when there exist } x \in M \text{ such that } s = xt \\ s \leqslant_{\mathcal{R}} t & \text{when there exist } y \in M \text{ such that } s = ty \end{array}$$

Clearly, $\leqslant_{\mathcal{J}}$, $\leqslant_{\mathcal{L}}$ and $\leqslant_{\mathcal{R}}$ are preorders (*i.e.*, they are reflexive and transitive). We write $<_{\mathcal{J}}$, $<_{\mathcal{L}}$ and $<_{\mathcal{R}}$ for their strict variants (for example, $s <_{\mathcal{J}} t$ when $s \leqslant_{\mathcal{J}} t$ but $t \not\leqslant_{\mathcal{J}} s$). Finally, we write $\mathcal{J}$, $\mathcal{L}$ and $\mathcal{R}$ for the corresponding equivalence relations (for example, $s \mathcal{J} t$ when $s \leqslant_{\mathcal{J}} t$ and $t \leqslant_{\mathcal{J}} s$). There are many technical results about Green relations. We shall only need the following simple lemma which applies to *finite* monoids (we recall its proof in appendix).

▶ **Lemma 1.** *Consider a finite monoid M and $s, t \in M$ such that $s \mathcal{J} t$. Then, $s \leqslant_{\mathcal{R}} t$ implies $s \mathcal{R} t$. Symmetrically, $s \leqslant_{\mathcal{L}} t$ implies $s \mathcal{L} t$.*

Observe that A^* is a monoid whose multiplication is concatenation (the neutral element is ε). 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 α when there exists a set $F \subseteq M$ such that $L = \alpha^{-1}(F)$.

Given any language L, there exists a canonical morphism which recognizes it. 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 $xuy \in L \Leftrightarrow xvy \in L$ 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 recognizing L called the *syntactic morphism of L*. Finally, 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 $\alpha_L : A^* \to M_L$ from any representation of L (such as a finite automaton).

Decision problems. The two problems that we consider in the paper are both parametrized by an arbitrary class of languages $\mathcal{C}$: they serve as mathematical tools for analyzing $\mathcal{C}$. The *$\mathcal{C}$-membership problem* is the simplest one. It takes as input a single regular language L and asks whether $L \in \mathcal{C}$. The second one, *$\mathcal{C}$-separation*, is more general: it takes **two** regular languages L_1, L_2 as input and asks whether L_1 is $\mathcal{C}$-separable from L_2, that is, whether there exists $K \in \mathcal{C}$ such that $L_1 \subseteq K$ and $L_2 \cap K = \emptyset$. The language K is called a *separator* of L_1 and L_2. Note that $\mathcal{C}$-membership is easily reduced to $\mathcal{C}$-separation: given any regular language L, we have $L \in \mathcal{C}$ if and only if L is $\mathcal{C}$-separable from $A^* \setminus L$ (which is also regular).

▶ Remark. When $\mathcal{C}$ is closed under complement (which is always the case in the paper), L_1 is $\mathcal{C}$-separable from L_2 if and only if L_2 is $\mathcal{C}$-separable from L_1.

3 Unambiguous polynomial closure

In this section, we define the unambiguous polynomial closure operation, which is the main focus of the paper. Furthermore, we investigate the associated membership problem.

3.1 Definition

Given two languages $H, L \subseteq A^*$, we say that their concatenation HL is *unambiguous* when any word $w \in HL$ admits a *unique* decomposition witnessing this membership: for any $u, u' \in H$ and $v, v' \in L$, if $w = uv = u'v'$, then $u = u'$ and $v = v'$. More generally, we say that a product of n languages $L_1 \cdots L_n$ is *unambiguous* when any word $w \in L_1 \cdots L_n$ admits a *unique* decomposition witnessing this membership. Note that unambiguous *marked* concatenations are well-defined: HaL is a product of three languages, namely H, $\{a\}$ and L.

▶ Remark. Clearly, not all products are unambiguous. For example, A^*aA^* is ambiguous: $aa \in A^*aA^*$ admits two decompositions witnessing this membership (εaa and $aa\varepsilon$).

▶ Remark. Being unambiguous is a *semantic* property: whether HL is unambiguous may not be apparent on the definitions of H and L. Moreover, this depends on the *product* HL and not only on the resulting language $K = HL$. It may happen that two products represent the same language but one is unambiguous while the other is not. For example, A^*aA^* is ambiguous while $(A \setminus \{a\})^*aA^*$ (which represents the same language) is unambiguous.

In the paper, we shall only need to use two special kinds of unambiguous products, which we now present. Let $K, L \subseteq A^*$ and $a \in A$. We say that the marked concatenation KaL,

- is *left deterministic* when $K \cap KaA^* = \emptyset$.
- is *right deterministic* when $L \cap A^*aL = \emptyset$.

▶ **Fact 2.** *Any left or right deterministic marked concatenation is unambiguous.*

We use these definitions to introduce three standard operations on classes of languages. Consider an arbitrary class $\mathcal{C}$.

- The *polynomial closure* of $\mathcal{C}$, denoted by $Pol(\mathcal{C})$, is the smallest class containing $\mathcal{C}$ and closed under marked concatenation and union: for any $H, L \in Pol(\mathcal{C})$ and $a \in A$, we have $HaL \in Pol(\mathcal{C})$ and $H \cup L \in Pol(\mathcal{C})$. Furthermore, we denote by *co-$Pol(\mathcal{C})$* the class containing all complements of languages in $Pol(\mathcal{C})$: $L \in$ *co-$Pol(\mathcal{C})$* when $A^* \setminus L \in Pol(\mathcal{C})$.

- The *unambiguous polynomial closure* of $\mathcal{C}$, denoted by $UPol(\mathcal{C})$, is the smallest class containing $\mathcal{C}$ and closed under *unambiguous* marked concatenation and *disjoint* union. That is, for any $H, L \in UPol(\mathcal{C})$ and $a \in A$, if HaL is unambiguous, then $HaL \in UPol(\mathcal{C})$ and if $H \cap L = \emptyset$, then $H \uplus L \in UPol(\mathcal{C})$. Here, we denote union by "$\uplus$" to underline the fact that H and L are disjoint (we use this convention in the whole paper).
- The *alternating deterministic closure* of $\mathcal{C}$, denoted by $ADet(\mathcal{C})$, is the smallest class containing $\mathcal{C}$ and closed under *deterministic* marked concatenation and *disjoint* union. That is, for any $H, L \in ADet(\mathcal{C})$ and $a \in A$, if HaL is either left or right deterministic, then $HaL \in ADet(\mathcal{C})$ and if $H \cap L = \emptyset$, then $H \uplus L \in ADet(\mathcal{C})$.

It is immediate by definition and Fact 2 that we have $\mathcal{C} \subseteq ADet(\mathcal{C}) \subseteq UPol(\mathcal{C}) \subseteq Pol(\mathcal{C})$. In general the inclusion $UPol(\mathcal{C}) \subseteq Pol(\mathcal{C})$ is strict. On the other hand, we shall prove that when $\mathcal{C}$ is a quotienting Boolean algebra, $ADet(\mathcal{C}) = UPol(\mathcal{C})$.

It is not immediate that $Pol(\mathcal{C})$, $UPol(\mathcal{C})$ and $ADet(\mathcal{C})$ have 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 these three classes as well. It was shown by Arfi [3] that when $\mathcal{C}$ is a quotienting lattice of regular languages, then $Pol(\mathcal{C})$ is one as well. Here, we are mostly interested in $UPol(\mathcal{C})$. We prove the following theorem which combines and extends several results by Pin, Straubing, Thérien and Weil [9, 11].

▶ **Theorem 3.** *Let $\mathcal{C}$ be a quotienting Boolean algebra of regular languages. Then, $UPol(\mathcal{C})$ is a quotienting Boolean algebra as well. Moreover, $UPol(\mathcal{C}) = ADet(\mathcal{C}) = Pol(\mathcal{C}) \cap co\text{-}Pol(\mathcal{C})$.*

That $UPol(\mathcal{C})$ is a quotienting Boolean algebra of regular languages is due to Pin, Straubing and Thérien [9]. The correspondence between $UPol(\mathcal{C})$ and $Pol(\mathcal{C}) \cap co\text{-}Pol(\mathcal{C})$ is due to Pin and Weil [11]. The correspondence between $UPol(\mathcal{C})$ and $ADet(\mathcal{C})$ is a new result, up to our knowledge. Let us point out that the original proofs of these results require a stronger hypothesis on $\mathcal{C}$, which needs additionally to be closed under inverse morphic image. Moreover, these proofs require to introduce and manipulate a lot of algebraic machinery. This is because they are based on a generic algebraic characterization of $UPol(\mathcal{C})$.

While we use a similar approach (*i.e.*, we prove a generic algebraic characterization of $UPol(\mathcal{C})$), our argument is much more elementary. The only algebraic notion that we need is the syntactic morphism of a regular language.

3.2 Algebraic characterization

We now present a generic algebraic characterization of $UPol(\mathcal{C})$. It holds provided that $\mathcal{C}$ is a quotienting Boolean algebra of regular languages. It implies Theorem 3, but also that $UPol(\mathcal{C})$-membership reduces to $\mathcal{C}$-membership.

The characterization is parameterized by two relations that we define now. Let $\mathcal{C}$ be some class of languages. Consider a finite monoid M and a *surjective* morphism $\alpha : A^* \to M$ (such as the syntactic morphism of some language). Given a pair $(s, t) \in M \times M$,

- (s, t) is a *$\mathcal{C}$-pair* (for α) when **no** language of $\mathcal{C}$ can separate $\alpha^{-1}(s)$ from $\alpha^{-1}(t)$.
- (s, t) is a *saturated $\mathcal{C}$-pair* (for α) when **no** language of $\mathcal{C}$ **recognized by α** can separate $\alpha^{-1}(s)$ from $\alpha^{-1}(t)$.

Note that any $\mathcal{C}$-pair is also a saturated $\mathcal{C}$-pair (the converse is not true in general). By definition, we are able to compute all $\mathcal{C}$-pairs as soon as we have an algorithm for $\mathcal{C}$-separation. On the other hand, computing all saturated $\mathcal{C}$-pairs boils down to deciding $\mathcal{C}$-membership, as it suffices to check which languages recognized by α (the potential separators) belong to $\mathcal{C}$.

▶ Remark. An equivalent definition of the saturated $\mathcal{C}$-pairs is to introduce them as the transitive closure of the $\mathcal{C}$-pairs. We prove this in appendix. In fact, when $\mathcal{C}$ is a quotienting Boolean algebra, the saturated $\mathcal{C}$-pair relation is a congruence whose equivalence classes correspond exactly to the languages recognized by α which belong to $\mathcal{C}$.

We may now state the following characterization of $UPol(\mathcal{C})$.

▶ **Theorem 4.** *Let $\mathcal{C}$ be a quotienting Boolean algebra of regular languages. Consider a regular language L and let $\alpha : A^* \to M$ be its syntactic morphism. The following are equivalent:*

1. *$L \in UPol(\mathcal{C})$.*
2. *$L \in ADet(\mathcal{C})$.*
3. *$L \in Pol(\mathcal{C}) \cap co\text{-}Pol(\mathcal{C})$.*
4. *For all $\mathcal{C}$-pairs $(s,t) \in M^2$, we have $s^{\omega+1} = s^\omega t s^\omega$.*
5. *For all saturated $\mathcal{C}$-pairs $(s,t) \in M^2$, we have $s^{\omega+1} = s^\omega t s^\omega$.*

Theorem 3 is a simple corollary of Theorem 4 (it is straightforward to verify that any class satisfying Item (4) in the theorem has to be a quotienting Boolean algebra). Another consequence is that if $\mathcal{C}$ is a quotienting Boolean algebra of regular languages, $UPol(\mathcal{C})$-membership reduces to the same problem for $\mathcal{C}$. Indeed, given as input a regular language L, one may compute its syntactic morphism α. By Theorem 4, deciding whether $L \in UPol(\mathcal{C})$ amounts to checking whether α satisfies Item (5). This is possible provided that we have all saturated $\mathcal{C}$-pairs for α in hand. In turn, an algorithm for $\mathcal{C}$-membership immediately yields an algorithm for computing them all. Altogether, we obtain the following corollary.

▶ **Corollary 5.** *Let $\mathcal{C}$ be a quotienting Boolean algebra of regular languages and assume that $\mathcal{C}$-membership is decidable. Then $UPol(\mathcal{C})$-membership is decidable as well.*

We now focus on proving Theorem 4. A first point is that we do not show the equivalence (3) $\Leftrightarrow$ (4): it follows from the generic characterization of $Pol(\mathcal{C})$ which is not our main focus in the paper (a full proof is available in [15]). Here, we concentrate on proving the implications (1) $\Rightarrow$ (4) $\Rightarrow$ (5) $\Rightarrow$ (2) $\Rightarrow$ (1). The implication (2) $\Rightarrow$ (1) ($ADet(\mathcal{C}) \subseteq UPol(\mathcal{C})$) is immediate. Even though the presentation is different, the equivalence (4) $\Leftrightarrow$ (5) is a result of [1] (which investigates $Pol(\mathcal{C}) \cap co\text{-}Pol(\mathcal{C})$). We prove this equivalence in appendix. We postpone the implication (1) $\Rightarrow$ (4) to the appendix as well, to focus on (5) $\Rightarrow$ (2), which is the most interesting implication: when a language satisfies (5), we show that it belongs to $ADet(\mathcal{C})$.

We fix a quotienting Boolean algebra of regular languages $\mathcal{C}$ for the proof. Consider an arbitrary surjective morphism $\alpha : A^* \to M$ satisfying Item (5) in Theorem 4. We show that any language recognized by α belongs to $ADet(\mathcal{C})$. We start with a preliminary lemma.

▶ **Lemma 6.** *There exists a finite monoid N and a surjective morphism $\beta : M \to N$ which satisfies the following properties:*

- *For any $s,t \in M$, (s,t) is a saturated $\mathcal{C}$-pair if and only if $\beta(s) = \beta(t)$.*
- *Any language recognized by the composition $\gamma = \beta \circ \alpha : A^* \to N$ belongs to $\mathcal{C}$.*

Lemma 6 is obtained by proving that the saturated $\mathcal{C}$-pair relation is a congruence on M and that for any equivalence class $F \subseteq M$, $\alpha^{-1}(F) \in \mathcal{C}$. It then suffices to define N as the quotient of M by this congruence. The proof is presented in appendix.

Let us come back to the main proof. Let $\beta : M \to N$ and the composition $\gamma = \beta \circ \alpha$ be defined as in Lemma 6. Given any $r_1, r_2, s \in M$ and any $x \in N$, we define:

$$L_s^x[r_1, r_2] = \{w \in \gamma^{-1}(x) \mid r_1 \cdot \alpha(w) \cdot r_2 = s\}.$$

The purpose of introducing $L_s^x[r_1, r_2]$ is that it provides induction parameters s, r_1, r_2 and it coincides with $\alpha^{-1}(s)$ when $x = \beta(s)$, $r_1 = r_2 = 1_M$. Our goal is to show that it is in $ADet(\mathcal{C})$.

▶ **Proposition 7.** *Let $r_1, r_2, s \in M$ and $x \in N$. Then, $L_s^x[r_1, r_2] \in ADet(\mathcal{C})$.*

Before proving this proposition, let us use it to finish the main proof. By definition, a language recognized by α is a disjoint union of sets $\alpha^{-1}(s)$ for $s \in M$. Therefore, it suffices to prove that $\alpha^{-1}(s) \in ADet(\mathcal{C})$ for any $s \in M$. Let $x = \beta(s)$. Clearly, $L_s^{\beta(s)}[1_M, 1_M] = \alpha^{-1}(s)$. Thus, Proposition 7 yields that $\alpha^{-1}(s) \in ADet(\mathcal{C})$, finishing the proof.

It remains to prove Proposition 7. We let $r_1, r_2, s \in M$ and $x \in N$. Our objective is to show that $L_s^x[r_1, r_2] \in ADet(\mathcal{C})$. Observe that we may assume without loss of generality that $\beta(s) = \beta(r_1)x\beta(r_2)$. Otherwise, $L_s^x[r_1, r_2] = \emptyset \in ADet(\mathcal{C})$ by definition and the result is immediate. The proof is an induction on the three following parameters listed by order of importance (the three of them depend on Green's relations in both M and N):

1. The *rank of* $\beta(s)$ which is the number of elements $y \in N$ such that $\beta(s) \leqslant_{\mathcal{J}} y$.
2. The *right index of* r_1 which is the number of elements $t \in M$ such that $t \leqslant_{\mathcal{R}} r_1$.
3. The *left index of* r_2 which is the number of elements $t \in M$ such that $t \leqslant_{\mathcal{L}} r_2$.

We consider three cases depending on the following properties of s, r_1, r_2 and x.

- We say that x is *smooth* when $x \mathrel{\mathcal{J}} \beta(s)$.
- We say that r_1 is *right stable* when there exists $t \in M$ such that $\beta(t) \mathrel{\mathcal{R}} x$ and $r_1 t \mathrel{\mathcal{R}} r_1$.
- We say that r_2 is *left stable* when there exists $t \in M$ such that $\beta(t) \mathrel{\mathcal{L}} x$ and $t r_2 \mathrel{\mathcal{L}} r_2$.

In the base case, we assume that all three properties hold. Otherwise, we consider two inductive cases. First, we assume that x is not smooth. Then, we assume that either r_1 is not right stable or r_2 is not left stable.

Base case. Assume that x is smooth and that r_1, r_2 are respectively right and left stable. We use this hypothesis to prove the following lemma.

▶ **Lemma 8.** *For any $u, v \in \gamma^{-1}(x)$, we have $r_1\alpha(u)r_2 = r_1\alpha(v)r_2$.*

Observe that Lemma 8 concludes the proof. Indeed, by definition of $L_s^x[r_1, r_2]$, it implies that either $L_s^x[r_1, r_2] = \gamma^{-1}(x)$ (when $r_1\alpha(w)r_2 = s$ for all $w \in \gamma^{-1}(x)$) or $L_s^x[r_1, r_2] = \emptyset$ (when $r_1\alpha(w)r_2 \neq s$ for all $w \in \gamma^{-1}(x)$). Since both of these languages belong to $\mathcal{C} \subseteq ADet(\mathcal{C})$ by Lemma 6, Proposition 7 follows. It remains to prove Lemma 8 to conclude the base case. The argument relies on the following fact (this is where we use our hypothesis on r_1 and r_2).

▶ **Fact 9.** *When Item (5) in Theorem 4 holds, the two following properties hold as well:*

- *For all $t \in M$ such that $\beta(t) \mathrel{\mathcal{R}} x$, we have $r_1 t \mathrel{\mathcal{R}} r_1$.*
- *For all $t \in M$ such that $\beta(t) \mathrel{\mathcal{L}} x$, we have $t r_2 \mathrel{\mathcal{L}} r_2$.*

Let us first use the fact to prove Lemma 8 and finish the base case. Consider $u, v \in \gamma^{-1}(x)$, *i.e.*, $\beta(\alpha(u)) = \beta(\alpha(v)) = x$. We show that $r_1\alpha(u)r_2 = r_1\alpha(v)r_2$.

By hypothesis, we have $\beta(s) = \beta(r_1)x\beta(r_2)$. Moreover, $\beta(s) \mathrel{\mathcal{J}} x$ since x is smooth by hypothesis. Thus, $x\beta(r_2) \mathrel{\mathcal{J}} x$ and $\beta(r_1)x \mathrel{\mathcal{J}} x$. Hence, since $x\beta(r_2) \leqslant_{\mathcal{R}} x$ and $\beta(r_1)x \leqslant_{\mathcal{L}} x$, Lemma 1 implies $x\beta(r_2) \mathrel{\mathcal{R}} x$ and $\beta(r_1)x \mathrel{\mathcal{L}} x$. Since $\beta(\alpha(u)) = x$, this yields $\beta(\alpha(u)r_2) \mathrel{\mathcal{R}} x$ and $\beta(r_1\alpha(u)) \mathrel{\mathcal{L}} x$. By Fact 9, it follows that $r_1\alpha(u)r_2 \mathrel{\mathcal{R}} r_1$ and $r_1\alpha(u)r_2 \mathrel{\mathcal{L}} r_2$. We get $p, q \in M$ such that $r_1 = r_1\alpha(u)r_2p$ and $r_2 = qr_1\alpha(u)r_2$. Let $t = qr_1\alpha(u)r_2p = r_2p = qr_1$. We combine our two equalities for r_1 and r_2 to obtain,

$$r_1 = r_1\alpha(u)t = r_1(\alpha(u)t)^\omega \quad \text{and} \quad r_2 = t\alpha(u)r_2 = (t\alpha(u))^{\omega+1}r_2.$$

Since $\beta(\alpha(u)) = \beta(\alpha(v))$, we know that $\beta(\alpha(u)t) = \beta(\alpha(v)t)$. Therefore, $(\alpha(u)t, \alpha(v)t)$ is a saturated $\mathcal{C}$-pair by Lemma 6, and Item (5) yields $(\alpha(u)t)^{\omega+1} = (\alpha(u)t)^{\omega}\alpha(v)t(\alpha(u)t)^{\omega}$. We may now multiply by r_1 on the left and by $\alpha(u)r_2$ on the right to get,

$$r_1(\alpha(u)t)^{\omega}\alpha(u)(t\alpha(u))^{\omega+1}r_2 = r_1(\alpha(u)t)^{\omega}\alpha(v)(t\alpha(u))^{\omega+1}r_2.$$

Since we already established that $r_1 = r_1(\alpha(u)t)^{\omega}$ and $r_2 = (t\alpha(u))^{\omega+1}r_2$, we get as desired that $r_1\alpha(u)r_2 = r_1\alpha(v)r_2$, finishing the proof of Lemma 8. It remains to prove Fact 9.

Proof of Fact 9. By symmetry, we focus on the first property and leave the second to the reader. Let $t \in M$ such that $\beta(t)\ \mathcal{R}\ x$. We show that $r_1 t\ \mathcal{R}\ r_1$. By hypothesis, r_1 is right stable which yields $t' \in M$ such that $\beta(t')\ \mathcal{R}\ x\ \mathcal{R}\ \beta(t)$ and $r_1 t'\ \mathcal{R}\ r_1$. Since $\beta(t')\ \mathcal{R}\ \beta(t)$, we have $y \in N$ such that $\beta(t') = \beta(t)y$. Let $p \in M$ such that $\beta(p) = y$: we have $\beta(t') = \beta(tp)$. Since $r_1 t'\ \mathcal{R}\ r_1$, we have $q \in M$ such that $r_1 = r_1 t' q$ which yields $r_1 = r_1(t'q)^{\omega} = r_1(t'q)^{\omega+1}$. We have $\beta(t'q) = \beta(tpq)$ which means that $(t'q, tpq)$ is a saturated $\mathcal{C}$-pair by Lemma 6. Therefore, Equation (5) yields that $(t'q)^{\omega+1} = (t'q)^{\omega}tpq(t'q)^{\omega}$. Finally, we obtain,

$$r_1 = r_1(t'q)^{\omega+1} = r_1(t'q)^{\omega}tpq(t'q)^{\omega} = r_1tpq(t'q)^{\omega}.$$

This implies that $r_1 \leqslant_{\mathcal{R}} r_1 t$. Since it is immediate that $r_1 t \leqslant_{\mathcal{R}} r_1$, we get $r_1 t\ \mathcal{R}\ r_1$. ◀

First inductive case. We now assume that x is not smooth: x and $\beta(s)$ are not $\mathcal{J}$-equivalent. We use induction on our first parameter (the rank of $\beta(s)$). Recall that we assumed $\beta(s) = \beta(r_1)x\beta(r_2)$, which yields $\beta(s) \leqslant_{\mathcal{J}} x$. Thus, we have $\beta(s) <_{\mathcal{J}} x$ by hypothesis.

By definition, $L_s^x[r_1, r_2]$ is the disjoint union of all languages $\alpha^{-1}(t)$ where $t \in M$ satisfies $\beta(t) = x$ and $r_1 t r_2 = s$. Therefore, it suffices to show that for any $t \in M$ such that $\beta(t) = x$, we have $\alpha^{-1}(t) \in ADet(\mathcal{C})$. This is immediate by induction. Indeed, since $\beta(t) = x$, we have $\alpha^{-1}(t) = L_t^x[1_M, 1_M]$. Moreover, since $\beta(s) <_{\mathcal{J}} x$, we have $\beta(s) <_{\mathcal{J}} \beta(t)$. It follows that the rank of $\beta(t)$ is strictly smaller than the one of $\beta(s)$. Hence, we may apply induction on our first and most important parameter to get $L_t^x[1_M, 1_M] \in ADet(\mathcal{C})$.

Second inductive case. We assume that either r_1 is not *right stable* or r_2 is not *left stable*. By symmetry, we treat the case when r_1 is not right stable and leave the other to the reader.

▶ Remark. We only apply induction on our two first parameters. Moreover, we show that $L_s^x[r_1, r_2]$ is built from languages in $ADet(\mathcal{C})$ (obtained from induction) using only disjoint union and left deterministic marked concatenations. Induction on our third parameter and right deterministic marked concatenations are used in the case when r_2 is not left stable.

Observe that we have $x <_{\mathcal{J}} 1_N$ (x is not maximal for $\leqslant_{\mathcal{J}}$). Indeed, otherwise, we would have $x\ \mathcal{R}\ 1_N$ by Lemma 1 and r_1 would be left stable: $1_M \in M$ would satisfy $\beta(1_M) = 1_N\ \mathcal{R}\ x$ and $r_1 \cdot 1_M = r_1\ \mathcal{R}\ r_1$. Therefore, there are elements $y \in N$ such that $x <_{\mathcal{J}} y$.

We use this observation to define T as the set of all triples $(y, a, z) \in N \times A \times N$ such that $x = y \cdot \gamma(a) \cdot z$, $x <_{\mathcal{J}} y$ and $x\ \mathcal{J}\ y \cdot \gamma(a)$. Using the definition of T and the fact that $x <_{\mathcal{J}} 1_N$, one may decompose $L_s^x[r_1, r_2]$ as follows (this lemma is proved in appendix).

▶ **Lemma 10.** *The language $L_s^x[r_1, r_2]$ is equal to the following disjoint union,*

$$L_s^x[r_1, r_2] = \biguplus_{(y,a,z)\in T} \left(\biguplus_{t\in\beta^{-1}(y)} \alpha^{-1}(t) \cdot a \cdot L_s^z[r_1 t\alpha(a), r_2] \right).$$

We now use Lemma 10 to show as desired that $L_s^x[r_1, r_2] \in ADet(\mathcal{C})$. Since $ADet(\mathcal{C})$ is closed under disjoint union by definition, it suffices to show that for any $(y, a, z) \in T$ and any $t \in \beta^{-1}(y)$, we have,

$$\alpha^{-1}(t) \cdot a \cdot L_s^z[r_1 t\alpha(a), r_2] \in ADet(\mathcal{C}).$$

We prove that this is a left deterministic marked concatenation of two languages in $ADet(\mathcal{C})$ which concludes the proof.

We start with $\alpha^{-1}(t) \in ADet(\mathcal{C})$. Since $y = \beta(t)$, we have $\alpha^{-1}(t) = L_t^y[1_M, 1_M]$. Moreover, since $\beta(s) = \beta(r_1)x\beta(r_2)$, we have $\beta(s) \leqslant_{\mathcal{J}} x$. Finally, by definition of T we have $x <_{\mathcal{J}} y = \beta(t)$. Altogether, we get $\beta(s) <_{\mathcal{J}} \beta(t)$: the rank of $\beta(t)$ is strictly smaller than the one of $\beta(s)$ and induction on our first parameter yields $\alpha^{-1}(t) = L_t^y[1_M, 1_M] \in ADet(\mathcal{C})$.

We turn to $L_s^z[r_1 t\alpha(a), r_2] \in ADet(\mathcal{C})$. By definition of T, we have $x \mathcal{J} y \cdot \gamma(a)$ and $x = y \cdot \gamma(a) \cdot z$ which yields that $x \mathcal{R} y \cdot \gamma(a)$ by Lemma 1. Moreover, since $y = \beta(t)$, it follows that $x \mathcal{R} \beta(t\alpha(a))$. Therefore, since we know that r_1 is **not** right stable (this is our hypothesis), it follows that r_1 and $r_1 t\alpha(a)$ are not $\mathcal{R}$-equivalent. Since it is clear that $r_1 t\alpha(a) \leqslant_{\mathcal{R}} r_1$, it follows that $r_1 t\alpha(a) <_{\mathcal{R}} r_1$: the right index of $r_1 t\alpha(a)$ is strictly smaller than the one of r_1. By induction on our second parameter, we then get that $L_s^z[r_1 t\alpha(a), r_2] \in ADet(\mathcal{C})$.

It remains to show that $\alpha^{-1}(t) \cdot a \cdot L_s^z[r_1 t\alpha(a), r_2]$ is a left deterministic marked concatenation, *i.e.*, that $\alpha^{-1}(t) \cap \alpha^{-1}(t)aA^* = \emptyset$. Since $\beta(t) = y$, we have $\alpha^{-1}(t) \subseteq \gamma^{-1}(y)$ and it suffices to show that $\gamma^{-1}(y) \cap \gamma^{-1}(y)aA^* = \emptyset$. Let $w \in \gamma^{-1}(y)$ and $w' \in \gamma^{-1}(y)aA^*$, we show that $w \neq w'$. Since $(x, a, z) \in T$, we have $x <_{\mathcal{J}} y$ and $x \mathcal{J} y \cdot \gamma(a)$. It follows that $y \cdot \gamma(a) <_{\mathcal{J}} y$. Finally, we have $\gamma(w) = y$ and $\gamma(w') = y\gamma(a)y'$ for some $y' \in N$. This implies that $\gamma(w') \leqslant_{\mathcal{J}} y\gamma(a) <_{\mathcal{J}} y = \gamma(w)$. Therefore $\gamma(w) \neq \gamma(w')$ which implies that $w \neq w'$.

4 Separation

We now turn to separation for $UPol(\mathcal{C})$ and show that the problem is decidable for any finite quotienting Boolean algebra $\mathcal{C}$. For the sake of avoiding clutter, we fix $\mathcal{C}$ for the section.

▶ Remark. This result may seem weak: our solution for $UPol(\mathcal{C})$-separation requires $\mathcal{C}$ to be finite while $UPol(\mathcal{C})$-membership reduces to $\mathcal{C}$-membership. This intuition is wrong: the result on separation is the strongest. The proof of Theorem 4 shows that when $L \in UPol(\mathcal{C})$, the basic languages in $\mathcal{C}$ needed to build L are all recognized by the syntactic morphism of L. Hence, $L \in UPol(\mathcal{C})$ if and only if $L \in UPol(\mathcal{D})$ where $\mathcal{D} \subseteq \mathcal{C}$ is a *finite class* obtained from the syntactic morphism of L. We lose this when moving to separation: the languages in $\mathcal{C}$ needed to build a potential separator in $UPol(\mathcal{C})$ may not be encoded in our two inputs.

Our algorithm is based on a general framework designed to handle separation problems and to present solutions in an elegant way. It was introduced in [16, 17]. We first summarize what we need in this framework to present our solution for $UPol(\mathcal{C})$-separation.

▶ Remark. The framework of [16, 17] is actually designed to handle a more general decision problem: covering, which generalizes separation to arbitrarily many input languages. Thus, our solution actually yields an algorithm for $UPol(\mathcal{C})$-covering as well. While we do not detail this point due to lack of space, this follows from the definitions of [16, 17].

4.1 Methodology

We briefly recall the framework of [16, 17]. We refer the reader to [17] for details. The approach is based on "rating maps", a notion designed to measure how well a language separate others.

The definition of rating maps relies on semirings. A *semiring* is a set R equipped with two binary operations $+$ and $\cdot$, called addition and multiplication, satisfying the following axioms:

- $(R, +)$ is a commutative monoid whose neutral element is denoted by 0_R.
- $(R, \cdot)$ is a monoid whose neutral element is denoted by 1_R.
- The multiplication distributes over addition: $r \cdot (s + t) = rs + rt$ and $(s + t) \cdot r = sr + tr$.
- The element 0_R is a zero for multiplication: for any $r \in R$, $0_R \cdot r = r \cdot 0_R = 0_R$.

Moreover, we say that a semiring R is *idempotent* when any element $r \in R$ is idempotent for addition: $r + r = r$. Any idempotent semiring R can be equipped with a canonical order "$\leq$": given $s, r \in R$, we have $s \leq r$ when $s + r = r$. It can be verified that this is indeed an order which is compatible with addition and multiplication (R being idempotent is required).

▶ **Example 11.** The set 2^{A^*} of all languages over A is an idempotent semiring: the addition is union and the multiplication is language concatenation. In this case, the canonical order is inclusion ($H \subseteq L$ if and only if $H \cup L = L$). Another important example is the powerset 2^M of any monoid M. Again the addition is union (thus, the order is inclusion). The multiplication is obtained from the one of M: given $S, T \in 2^M$, $S \cdot T = \{st \mid s \in S \text{ and } t \in T\}$.

Rating maps. A rating map[1] is a *semiring morphism*, $\rho : 2^{A^*} \to R$ where R is a *finite idempotent semiring*. It can be verified that any rating map is compatible with the canonical order ($K \subseteq L \Rightarrow \rho(K) \leq \rho(L)$). For the sake of improved readability, when applying a rating map ρ to a singleton language $\{w\}$, we shall simply write $\rho(w)$ for $\rho(\{w\})$. The connection with separation only requires to consider special rating maps called "*nice*". A rating map $\rho : 2^{A^*} \to R$ is *nice* when for any language $K \subseteq A^*$, $\rho(K) = \sum_{w \in K} \rho(w)$ (while infinite, this sum boils down to a finite one as R is a finite idempotent commutative monoid for addition).

▶ Remark. Any nice rating map $\rho : 2^{A^*} \to R$ is finitely representable: it is determined by the images $\rho(a)$ of letters $a \in A$. We may speak of algorithms whose inputs are nice rating maps.

Solving $UPol(\mathcal{C})$-separation requires to consider a special class of rating maps: the *$\mathcal{C}$-compatible ones* (our algorithm is restricted to them). The definition is based on a canonical equivalence $\sim_{\mathcal{C}}$ on A^* associated to $\mathcal{C}$. Given $u, v \in A^*$, we write $u \sim_{\mathcal{C}} v$ if and only if $u \in L \;\Leftrightarrow\; v \in L$ for all $L \in \mathcal{C}$. Clearly, $\sim_{\mathcal{C}}$ is an equivalence relation. For any word $w \in A^*$, we shall write $[w]_{\mathcal{C}} \subseteq A^*$ for the $\sim_{\mathcal{C}}$-class of w. Moreover, since $\mathcal{C}$ is a finite quotienting Boolean algebra, we have the following classical properties (we present a proof in appendix).

▶ **Lemma 12.** *The equivalence $\sim_{\mathcal{C}}$ is a congruence of finite index for word concatenation. Moreover, for any language $L \subseteq A^*$, we have $L \in \mathcal{C}$ if and only if L is a union of $\sim_{\mathcal{C}}$-classes.*

Lemma 12 implies that the set $A^*/\sim_{\mathcal{C}}$ of $\sim_{\mathcal{C}}$-classes is a finite monoid and the map $w \mapsto [w]_{\mathcal{C}}$ is a morphism. For the sake of avoiding confusion with language concatenation, we shall write "$\bullet$" for the monoid multiplication of $A^*/\sim_{\mathcal{C}}$. In general, if $C, D \subseteq A^*$ are $\sim_{\mathcal{C}}$-classes, then $C \bullet D \neq CD$ (usually, CD is not even a $\sim_{\mathcal{C}}$-class).

We may now define $\mathcal{C}$-compatibility. We say that a rating map $\rho : 2^{A^*} \to R$ is *$\mathcal{C}$-compatible* when for any two $\sim_{\mathcal{C}}$-classes C and D, if there exists an element $r \in R \setminus \{0_R\}$ such that $r \leq \rho(C)$ and $r \leq \rho(D)$, then $C = D$.

[1] What we call rating map here is called **multiplicative** rating map in [17] (the "true" rating maps are weaker and do not require a multiplication). We abuse terminology for the sake of improved readability.

Optimal covers. We use rating maps to define objects called "optimal universal $\mathcal{D}$-covers", which encode separation-related information. Here, $\mathcal{D}$ is an arbitrary fixed Boolean algebra for which one wants a $\mathcal{D}$-separation algorithm (we are interested in the case $\mathcal{D} = \mathit{UPol}(\mathcal{C})$). We fix $\mathcal{D}$ for the definition.

A *cover of some language* L is a *finite* set of languages $\mathbf{K}$ such that $L \subseteq \bigcup_{K \in \mathbf{K}} K$. When $L = A^*$, we speak of *universal cover*. Moreover, we say that $\mathbf{K}$ is a $\mathcal{D}$-cover when all $K \in \mathbf{K}$ belong to $\mathcal{D}$. A fixed rating map $\rho : 2^{A^*} \to R$ is used to define a "quality measure" for $\mathcal{D}$-covers which yields a notion of "best" universal $\mathcal{D}$-cover. Given a finite set of languages $\mathbf{K}$ (such as a universal $\mathcal{D}$-cover), the *ρ-imprint* $\mathcal{I}[\rho](\mathbf{K})$ of $\mathbf{K}$ is the following subset of R:

$$\mathcal{I}[\rho](\mathbf{K}) = \{r \in R \mid r \leq \rho(K) \text{ for some } K \in \mathbf{K}\}.$$

We now define the optimal universal $\mathcal{D}$-covers as those with the smallest possible ρ-imprint (with respect to inclusion). A *universal $\mathcal{D}$-cover $\mathbf{K}$ is optimal for ρ* when $\mathcal{I}[\rho](\mathbf{K}) \subseteq \mathcal{I}[\rho](\mathbf{K}')$ for any universal $\mathcal{D}$-cover $\mathbf{K}'$. In general, there can be infinitely many optimal universal $\mathcal{D}$-covers for a given rating map ρ. The crucial point is that there always exists a least one. This is simple and proved in [17]. The key idea is that there are finitely many possible ρ-imprints (since R is finite) and given two universal $\mathcal{D}$-covers, one may always build a third one which has a smaller ρ-imprint than the first two, by simple use of language intersections.

Finally, a key observation is that by definition, all optimal universal $\mathcal{D}$-covers for ρ share the same ρ-imprint. This unique ρ-imprint is a *canonical* object for $\mathcal{D}$ and ρ called the *$\mathcal{D}$-optimal universal ρ-imprint* and we denote it by $\mathcal{I}_{\mathcal{D}}[\rho]$. That is, $\mathcal{I}_{\mathcal{D}}[\rho] = \mathcal{I}[\rho](\mathbf{K})$ for any optimal universal $\mathcal{D}$-cover $\mathbf{K}$ for ρ.

The connection with separation. We may now explain how these notions are used to handle separation. This is summarized by the following lemma.

▶ **Lemma 13.** *Let $\mathcal{D}$ be a Boolean algebra and assume that there exists an algorithm that takes as input a nice $\mathcal{C}$-compatible rating map $\rho : 2^{A^*} \to R$ and outputs $\mathcal{I}_{\mathcal{D}}[\rho]$. Then, $\mathcal{D}$-separation is decidable.*

Let us sketch how to go from computing $\mathcal{D}$-optimal ρ-imprints to $\mathcal{D}$-separation (see [16] or the appendix for a full proof of Lemma 13). Consider two regular languages L_1 and L_2: we wish to know whether L_1 is $\mathcal{D}$-separable from L_2. Since $\mathcal{C}$ is finite, one can build a monoid morphism $\alpha : A^* \to M$, with M finite, recognizing both L_1 and L_2 as well as all languages in $\mathcal{C}$. Furthermore, one may lift α as a map $\rho : 2^{A^*} \to 2^M$ by defining $\rho(K) = \{\alpha(w) \mid w \in K\}$ for any language $K \subseteq A^*$. It is simple to verify that this map ρ is a nice $\mathcal{C}$-compatible rating map. Moreover, the two following properties (which we prove in appendix) hold:

- L_1 is $\mathcal{D}$-separable from L_2 iff for any $s_1 \in \alpha(L_1)$ and $s_2 \in \alpha(L_2)$, we have $\{s_1, s_2\} \notin \mathcal{I}_{\mathcal{D}}[\rho]$.
- When the first item holds, one may build a separator in $\mathcal{D}$ from any optimal universal $\mathcal{D}$-cover $\mathbf{K}$ for ρ: this separator is the union of all languages intersecting L_1 in $\mathbf{K}$.

By the first item, having an algorithm that computes $\mathcal{I}_{\mathcal{D}}[\rho] \subseteq 2^M$ suffices to decide whether L_1 is $\mathcal{D}$-separable from L_2. Moreover, by the second item, having an algorithm that computes an optimal universal $\mathcal{D}$-cover $\mathbf{K}$ for ρ is enough to build a separator (when it exists).

▶ Remark. Here, we only use the sets of size two in $\mathcal{I}_{\mathcal{D}}[\rho] \subseteq 2^M$. However, $\mathcal{I}_{\mathcal{D}}[\rho]$ contains more information corresponding to the more general $\mathcal{D}$-covering problem considered in [16, 17].

4.2 Computing $\mathit{UPol}(\mathcal{C})$-optimal universal imprints

We use the framework defined above to present an algorithm for $\mathit{UPol}(\mathcal{C})$-separation. We give a characterization $\mathit{UPol}(\mathcal{C})$-optimal imprints. It yields a procedure for computing them.

Consider a rating map $\rho : 2^{A^*} \to R$. For any subset $S \subseteq R$, we say that S is *UPol*($\mathcal{C}$)*-saturated* (for ρ) if it contains the set $\mathcal{I}_{triv}[\rho] = \{r \in R \mid r \leq \rho(w) \text{ for some } w \in A^*\}$ and is closed under the following operations:

1. *Downset*: for any $s \in S$, if $r \in R$ satisfies $r \leq s$, then we have $r \in S$.
2. *Multiplication*: For any $s, t \in S$, we have $st \in S$.
3. *UPol*($\mathcal{C}$)*-closure*: Given two $\sim_{\mathcal{C}}$-classes C, D and $s, t \in S$ such that $s \leq \rho(C \bullet D)$ and $t \leq \rho(D \bullet C)$, we have $s^\omega \cdot \rho(C) \cdot t^\omega \in S$

We are ready to state the main theorem of this section: when ρ is $\mathcal{C}$-compatible, *UPol*($\mathcal{C}$)-saturation characterizes the *UPol*($\mathcal{C}$)-optimal universal ρ-imprint.

▶ **Theorem 14.** *Let $\rho : 2^{A^*} \to R$ be a $\mathcal{C}$-compatible rating map. Then, $\mathcal{I}_{UPol(\mathcal{C})}[\rho]$ is the smallest UPol($\mathcal{C}$)-saturated subset of R (with respect to inclusion).*

Clearly, given a nice $\mathcal{C}$-compatible rating map $\rho : 2^{A^*} \to R$ as input, one may compute the smallest *UPol*($\mathcal{C}$)-saturated subset of R with a least fixpoint algorithm. One starts from $\mathcal{I}_{triv}[\rho]$ (which is clearly computable) and saturates this set with the three above operations. Thus, we get a procedure for computing $\mathcal{I}_{UPol(\mathcal{C})}[\rho]$ from any input nice $\mathcal{C}$-compatible rating map. By Lemma 13, this yields the desired corollary: *UPol*($\mathcal{C}$)-separation is decidable.

▶ **Corollary 15.** *For any finite quotienting Boolean algebra $\mathcal{C}$, UPol($\mathcal{C}$)-separation is decidable.*

The proof of Theorem 14 is a difficult generalization of the argument we used to show the algebraic characterization of *UPol*($\mathcal{C}$) (*i.e.*, Theorem 4). It is presented in appendix. An interesting byproduct of this proof is an algorithm which computes optimal universal *UPol*($\mathcal{C}$)-covers (and therefore *UPol*($\mathcal{C}$)-separators when they exist, as we explained above).

5 Conclusion

We presented a new, self-contained proof that for any quotienting Boolean algebra regular languages $\mathcal{C}$, membership for *UPol*($\mathcal{C}$) reduces to membership for $\mathcal{C}$. An interesting byproduct of this proof is that *UPol*($\mathcal{C}$) corresponds exactly to the class *ADet*($\mathcal{C}$), which is obtained by restricting the unambiguous marked concatenations to left or right deterministic ones. Moreover, we showed that when $\mathcal{C}$ is a finite quotienting Boolean algebra, *UPol*($\mathcal{C}$)-separation is decidable. This completes similar results of [18] for *Pol*($\mathcal{C}$) and *Bool*(*Pol*($\mathcal{C}$)) and of [13] for *Pol*(*Bool*(*Pol*($\mathcal{C}$))). These results raise several natural questions.

Historically, *UPol*($\mathcal{C}$) was investigated together with two weaker operations: left and right deterministic closures. The left (resp. right) deterministic closure of $\mathcal{C}$, is the smallest class containing $\mathcal{C}$ closed under disjoint union and left (resp. right) deterministic marked concatenation. Our results can be adapted to these two weaker operations. In both cases, membership reduces to $\mathcal{C}$-membership when $\mathcal{C}$ is a quotienting Boolean algebra of regular languages and separation is decidable when $\mathcal{C}$ is a finite quotienting Boolean algebra. In fact, these operations are simpler to handle than *UPol*($\mathcal{C}$). We leave this for further work.

Another question is whether our results can be pushed to classes built by combining unambiguous polynomial closure with other operations. A natural example is as follows. It is known [13] that *Pol*(*Bool*(*Pol*($\mathcal{C}$)))-separation is decidable when $\mathcal{C}$ is a finite quotienting Boolean algebra. Is this true as well for *UPol*(*Bool*(*Pol*($\mathcal{C}$)))? This seems difficult: the proof of [13] crucially exploits the fact that *Pol*(*Bool*(*Pol*($\mathcal{C}$))) is closed under concatenation (which is not the case for *UPol*(*Bool*(*Pol*($\mathcal{C}$)))) to handle the first polynomial closure.

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. Polynomial operations on rational languages. In *Proceedings of the 4th Annual Symposium on Theoretical Aspects of Computer Science*, STACS'87, pages 198–206, Berlin, Heidelberg, 1987. Springer-Verlag.

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

4 Volker Diekert, Paul Gastin, and Manfred Kufleitner. A survey on small fragments of first-order logic over finite words. *International Journal of Foundations of Computer Science (IJFCS)*, 19(3):513–548, June 2008.

5 Kousha Etessami, Moshe Y. Vardi, and Thomas Wilke. First-order logic with two variables and unary temporal logic. In *Proceedings of the 12th Annual IEEE Symposium on Logic in Computer Science*, LICS'97, pages 228–235, 1997.

6 Kousha Etessami, Moshe Y. Vardi, and Thomas Wilke. First-order logic with two variables and unary temporal logic. *Information and Computation*, 179(2):279–295, 2002.

7 James Alexander Green. On the structure of semigroups. *Annals of Mathematics*, 54(1):163–172, 1951.

8 Jean-Éric Pin. Propriétés syntactiques du produit non ambigu. In *Proceedings of the 7th International Colloquium on Automata, Languages and Programming*, ICALP'80, pages 483–499, 1980.

9 Jean-Éric Pin, Howard Straubing, and Denis Thérien. Locally trivial categories and unambiguous concatenation. *Journal of Pure and Applied Algebra*, 52(3):297 – 311, 1988.

10 Jean-Éric Pin and Pascal Weil. Polynomial closure and unambiguous product. In *Proceedings of the 22nd International Colloquium on Automata, Languages and Programming*, ICALP'95, pages 348–359, Berlin, Heidelberg, 1995. Springer-Verlag.

11 Jean-Éric Pin and Pascal Weil. Polynomial closure and unambiguous product. *Theory of Computing Systems*, 30(4):383–422, 1997.

12 Thomas Place. Separating regular languages with two quantifiers alternations. In *Proceedings of the 30th Annual ACM/IEEE Symposium on Logic in Computer Science, (LICS'15)*, pages 202–213. IEEE Computer Society, 2015.

13 Thomas Place. Separating regular languages with two quantifiers alternations. Unpublished, a preliminary version can be found at `https://arxiv.org/abs/1707.03295`, 2018.

14 Thomas Place, Lorijn van Rooijen, and Marc Zeitoun. Separating regular languages by piecewise testable and unambiguous languages. In *Proceedings of the 38th International Symposium on Mathematical Foundations of Computer Science*, MFCS'13, pages 729–740, Berlin, Heidelberg, 2013. Springer-Verlag.

15 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.

16 Thomas Place and Marc Zeitoun. The covering problem: A unified approach for investigating the expressive power of logics. In *Proceedings of the 41st International Symposium on Mathematical Foundations of Computer Science*, MFCS'16, pages 77:1–77:15, 2016.

17 Thomas Place and Marc Zeitoun. The covering problem. Unpublished, a preliminary version can be found at https://arxiv.org/abs/1707.03370, 2017.

18 Thomas Place and Marc Zeitoun. Separation for dot-depth two. In *Proceedings of the 32th Annual ACM/IEEE Symposium on Logic in Computer Science, (LICS'17)*, pages 202–213. IEEE Computer Society, 2017.

19 Thomas Place and Marc Zeitoun. A generic characterization of Pol($\mathcal{C}$). Manuscript, `http://www.labri.fr/perso/zeitoun/research/pdf/polc.pdf`, 2018.

20 John L. Rhodes. A homomorphism theorem for finite semigroups. *Mathematical systems theory*, 1:289–304, 1967.

21 M.P. Schützenberger. Sur le produit de concaténation non ambigu. *Semigroup Forum*, 13:47–75, 1976.

22 Imre Simon. Piecewise testable events. In *Proceedings of the 2nd GI Conference on Automata Theory and Formal Languages*, pages 214–222, Berlin, Heidelberg, 1975. Springer-Verlag.

23 Pascal Tesson and Denis Therien. Diamonds are forever: The variety DA. In *Semigroups, Algorithms, Automata and Languages*, pages 475–500. World Scientific, 2002.

24 Denis Thérien and Thomas Wilke. Over words, two variables are as powerful as one quantifier alternation. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, STOC'98, pages 234–240, New York, NY, USA, 1998. ACM.

A Proofs for Lemma 1 and Lemma 12

Here, we prove two independent lemmas that we shall use throughout the whole appendix. The first one states a simple property of Green relations while the second one applies to canonical equivalence associated to some finite quotienting Boolean algebra.

A.1 Lemma 1

▶ **Lemma 1.** *Consider a finite monoid M and $s,t \in M$ such that $s \mathcal{J} t$. Then, $s \leqslant_{\mathcal{R}} t$ implies $s \mathcal{R} t$. Symmetrically, $s \leqslant_{\mathcal{L}} t$ implies $s \mathcal{L} t$.*

Proof. Consider a finite monoid M and $s,t \in M$ such that $s \mathcal{J} t$. We have to show that $s \leqslant_{\mathcal{R}} t$ implies $s \mathcal{R} t$ and $s \leqslant_{\mathcal{L}} t$ implies $s \mathcal{L} t$. By symmetry we only prove the former.

Assume that $s \leqslant_{\mathcal{R}} t$. This yields $x \in M$ such that $s = tx$. Moreover, since $s \mathcal{J} t$, we have $y, z \in M$ such that $t = ysz$. Therefore, we get that,

$$t = ytxz = y^\omega t(xz)^\omega = y^\omega t(xz)^\omega (xz)^\omega = t(xz)^\omega$$

Therefore, $t = tx(zx)^{\omega-1}z = s(zx)^{\omega-1}z$. This exactly says that $t \leqslant_{\mathcal{R}} s$. Since we already have $s \leqslant_{\mathcal{R}} t$, this yields $s \mathcal{R} t$, as desired. ◀

A.2 Lemma 12

▶ **Lemma 12.** *For any quotienting Boolean algebra $\mathcal{C}$, the equivalence $\sim_{\mathcal{C}}$ is a congruence of finite index for word concatenation. Moreover, for any language $L \subseteq A^*$, we have $L \in \mathcal{C}$ if and only if L is a union of $\sim_{\mathcal{C}}$-classes.*

Proof. Recall that given any finite quotienting Boolean algebra $\mathcal{C}$, we associate a canonical equivalence $\sim_{\mathcal{C}}$ on A^* to $\mathcal{C}$. Given $u, v \in A^*$, we write $u \sim_{\mathcal{C}} v$ when $u \in L \Leftrightarrow v \in L$ for all $L \in \mathcal{C}$. Clearly, $\sim_{\mathcal{C}}$ is an equivalence relation. We need to prove that $\sim_{\mathcal{C}}$ is a congruence of finite index for word concatenation and for any language $L \subseteq A^*$, we have $L \in \mathcal{C}$ if and only if L is a union of $\sim_{\mathcal{C}}$-classes.

It is immediate that $\sim_{\mathcal{C}}$ has finite index, since $\mathcal{C}$ is a finite class. Indeed, by definition, the $\sim_{\mathcal{C}}$-class $[w]_{\mathcal{C}}$ of a word $w \in A^*$ only depends on which languages of $\mathcal{C}$ contain w. Therefore, there are at most $2^{|\mathcal{C}|}$ classes, where $|\mathcal{C}|$ denotes the number of languages in $\mathcal{C}$.

We now prove that $\mathcal{C}$ is a congruence. This is because $\mathcal{C}$ is closed under quotients. Let $u_1, v_1, u_2, v_2 \in A^*$ be words such that $u_1 \sim_{\mathcal{C}} v_1$ and $u_2 \sim_{\mathcal{C}} v_2$. We have to show that $u_1u_2 \sim_{\mathcal{C}} v_1v_2$: for any $L \in \mathcal{C}$, $u_1u_2 \in L \Leftrightarrow v_1v_2 \in L$. Thus, we fix $L \in \mathcal{C}$. By symmetry, we only prove the left to right implication. Assume that $u_1u_2 \in L$, we show that $v_1v_2 \in L$. Since $u_1u_2 \in L$, we have $u_1 \in L(u_2)^{-1}$. Moreover, since $\mathcal{C}$ is closed under quotients, we have $L(u_2)^{-1} \in \mathcal{C}$. Therefore, $u_1 \sim_{\mathcal{C}} v_1$ yields that $v_1 \in L(u_2)^{-1}$, *i.e.*, $v_1u_2 \in L$. Consequently, we have $u_2 \in (v_1)^{-1}L$. Again, since $\mathcal{C}$ is closed under quotients, we have $(v_1)^{-1}L \in \mathcal{C}$. Thus, $u_2 \sim_{\mathcal{C}} v_2$ yields that $v_2 \in (v_1)^{-1}L$, which exactly says that $v_1v_2 \in L$, finishing the proof.

It remains to show that the languages in $\mathcal{C}$ are the unions of $\sim_{\mathcal{C}}$-classes. This is because $\mathcal{C}$ is a Boolean algebra. Consider $L \subseteq A^*$. We have two implications to show. That any $L \in \mathcal{C}$ is a union of $\sim_{\mathcal{C}}$-classes is immediate from the definition of $\sim_{\mathcal{C}}$. Indeed, given $L \in \mathcal{C}$, if $u \in L$ and $u \sim_{\mathcal{C}} v$ for some $v \in A^*$, then $v \in L$ by definition. Therefore, we concentrate on the converse implication: any union of $\sim_{\mathcal{C}}$-classes belongs to $\mathcal{C}$. Since $\mathcal{C}$ is closed under union, it suffices to show that any $\sim_{\mathcal{C}}$-class belongs to $\mathcal{C}$. Consider $w \in A^*$, we show that

$[w]_{\mathcal{C}} \in \mathcal{C}$. By definition, it is immediate that,

$$[w]_{\mathcal{C}} = \left(\bigcap_{\{L \in \mathcal{C} \mid w \in L\}} L \right) \cap \left(\bigcap_{\{L \in \mathcal{C} \mid w \notin L\}} A^* \setminus L \right).$$

Hence, since $\mathcal{C}$ is a *finite* Boolean algebra, it is immediate that $[w]_{\mathcal{C}} \in \mathcal{C}$. ◀

B A characteristic property of $UPol(\mathcal{C})$

This second appendix presents a characteristic property of $UPol(\mathcal{C})$. We shall use it in two proofs: the direction (1) ⇒ (4) in Theorem 4 (the algebraic characterization of $UPol(\mathcal{C})$) and one direction in Theorem 14 (the characterization of $UPol(\mathcal{C})$-optimal imprints).

This property applies to the unambiguous polynomial closure of finite quotienting Boolean algebras. Recall that when $\mathcal{C}$ is a finite quotienting Boolean algebra, we associate a canonical equivalence $\sim_{\mathcal{C}}$ over A^* (two words are equivalent when they belong to the same languages in $\mathcal{C}$). Since $\mathcal{C}$ is closed under quotients, $\sim_{\mathcal{C}}$ must be a congruence of finite index for word concatenation by Lemma 12.

▶ **Proposition 13.** *Let $\mathcal{C}$ be a finite quotienting Boolean algebra and let A be a finite alphabet. Consider a language $L \subseteq A^*$ in $UPol(\mathcal{C})$. Then, there exists an integer $h \in \mathbb{N}$ such that for any $\ell \geq h$ and $u_1, u_2, v, v', x, y, z \in A^*$ satisfying $v \sim_{\mathcal{C}} v'$, $vz \sim_{\mathcal{C}} v'z \sim_{\mathcal{C}} u_1$ and $zv \sim_{\mathcal{C}} zv' \sim_{\mathcal{C}} u_2$, we have,*

$$x(u_1)^\ell v (u_2)^\ell y \in L \quad \text{if and only if} \quad x(u_1)^\ell v' (u_2)^\ell y \in L.$$

Before proving Proposition 13, we present a corollary which states a simpler version of the same property. This variant is the one that we shall use when proving the direction (1) ⇒ (4) in Theorem 4 (the full variant stated in Proposition 13 is required for Theorem 14).

▶ **Corollary 14.** *Let $\mathcal{C}$ be a finite quotienting Boolean algebra and let A be a finite alphabet. Consider a language $L \subseteq A^*$ in $UPol(\mathcal{C})$. There exists an integer $h \in \mathbb{N}$ such that for any $\ell \geq h$ and $u, v, x, y \in A^*$ satisfying $u \sim_{\mathcal{C}} v$, we have,*

$$xu^{2\ell+1}y \in L \quad \text{if and only if} \quad xu^\ell v u^\ell y \in L$$

Proof. This is the special case of Proposition 13 when $u_1 = u_2 = v'$ and $z = \varepsilon$. ◀

We now concentrate on proving Proposition 13. We fix the finite quotienting Boolean algebra $\mathcal{C}$ for the proof. Consider a language $L \subseteq A^*$ in $UPol(\mathcal{C})$. We first need to choose a natural number $h \in \mathbb{N}$ depending on L.

Recall that since $\sim_{\mathcal{C}}$ is a congruence of finite index, the set $A^*/\sim_{\mathcal{C}}$ of $\sim_{\mathcal{C}}$-classes if a finite monoid. Thus, it has an idempotent power ω. We shall denote it by $p \geq 1$ in the proof. Observe that since the map $w \mapsto [w]_{\mathcal{C}}$ is a morphism, we have the following fact by definition of p.

▶ **Fact 15.** *For any $w \in A^*$ and any $m, m' \geq 1$, $w^{pm} \sim_{\mathcal{C}} w^{pm'}$.*

Moreover, since L belongs to $UPol(\mathcal{C})$, it is built from languages in $\mathcal{C}$ using only disjoint union and unambiguous marked concatenation. It is simple to verify that these two operations commute. Hence, L is a finite disjoint union of *unambiguous products* having the form:

$$L_0 a_1 L_1 \cdots a_m L_m,$$

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

$$h = (2n+1) \times p.$$

It remains to show that h satisfies the desired property. Let $\ell \geq h$ and $u_1, u_2, v, v', x, y, z \in A^*$ such that $v \sim_{\mathcal{C}} v'$, $vz \sim_{\mathcal{C}} v'z \sim_{\mathcal{C}} u_1$ and $zv \sim_{\mathcal{C}} zv' \sim_{\mathcal{C}} u_2$. We have to show that,

$$x(u_1)^\ell v(u_2)^\ell y \in L \quad \text{if and only if} \quad x(u_1)^\ell v'(u_2)^\ell y \in L$$

Since our hypotheses on v and v' are symmetrical, we focus on the left to right direction (the converse one is obtained by symmetry). Therefore, we assume that $x(u_1)^\ell v(u_2)^\ell y \in L$ and show that $x(u_1)^\ell v'(u_2)^\ell y \in L$.

By hypothesis, we know that there exists a unambiguous product $L_0a_1L_1 \cdots a_mL_m \subseteq L$ with $a_1, \ldots, a_m \in A$, $L_0, \ldots, L_m \in \mathcal{C}$ and $m \leq n$ such that $x(u_1)^\ell v(u_2)^\ell y \in L_0a_1L_1 \cdots a_mL_m$. It follows that $x(u_1)^\ell v(u_2)^\ell y$ admits a unique decomposition,

$$x(u_1)^\ell v(u_2)^\ell y = w_0a_1w_1 \cdots a_mw_m$$

such that $w_i \in L_i$ for all $i \leq m$. The argument is based on the following lemma which is where we use our choice of h and the fact that $L_0a_1L_1 \cdots a_mL_m$ is unambiguous. It states that the central infix v in $x(u_1)^\ell v(u_2)^\ell y$ must be contained within one of the infixes w_i.

▶ **Lemma 16.** *There exist $i \leq m$ and $x_1, x_2 \in A^*$ such that,*

- $w_i = x_1vx_2$.
- $w_0a_1w_1 \cdots a_ix_1 = x(u_1)^\ell$.
- $x_2a_{i+1} \cdots a_mw_m = (u_2)^\ell y$.

Before we prove Lemma 16, let us use it to finish the main argument. Recall that by hypothesis $v \sim_{\mathcal{C}} v'$. Thus, since $\sim_{\mathcal{C}}$ is a congruence, we have $w_i = x_1vx_2 \sim_{\mathcal{C}} x_1v'x_2$. Moreover, since $w_i \in L_i$ which is a language of $\mathcal{C}$, it follows from the definition of $\sim_{\mathcal{C}}$ that $x_1v'x_2 \in L_i$. Therefore, since $w_j \in L_j$ for all j,

$$w_0a_1w_1 \cdots a_ix_1v'x_2a_{i+1} \cdots a_mw_m \in L_0a_1L_1 \cdots a_mL_m$$

By the last two items in Lemma 16, this exactly says that $x(u_1)^\ell v'(u_2)^\ell y \in L_0a_1L_1 \cdots a_mL_m$. Since we have $L_0a_1L_1 \cdots a_mL_m \subseteq L$ by definition, this implies that $x(u_1)^\ell v'(u_2)^\ell y \in L$, finishing the proof. It remains to prove Lemma 16.

Proof of Lemma 16. Recall that we have the following equality,

$$x(u_1)^\ell v(u_2)^\ell y = w_0a_1w_1 \cdots a_mw_m$$

Since $\ell \geq h = (2n+1) \times p \geq (2m+1) \times p$, it follows from the pigeon-hole principle that there exist $i \leq j \leq m$ such that,

- An infix $(u_1)^p$ of the prefix $x(u_1)^\ell$ is contained in w_i.
- An infix $(u_2)^p$ of the suffix $(u_2)^\ell y$ is contained in w_j.

We show that $i = j$. It will then be immediate that v is contained in $w_i = w_j$ as well and the lemma will follow. We proceed by contradiction. Assume that $i < j$ and let $K = a_{i+1}L_{i+1} \cdots a_{j-1}L_{j-1}a_j$. Observe that since $L_0a_1L_1 \cdots a_mL_m$ is unambiguous, the

sub-product L_iKL_j must be unambiguous as well. Using our hypotheses, we contradict this property by exhibiting a word $w \in L_iKL_j$ admitting two decompositions witnessing this membership.

We define $w' = a_{i+1}w_{i+1}\cdots a_{j-1}w_{j-1}a_j$. Note that $w' \in K$ by definition of K. By construction of i and j, we have words $x_1, y_1, x_2, y_2 \in A^*$ satisfying the following properties:

- $w_i = x_1(u_1)^p y_1$.
- $w_j = x_2(u_2)^p y_2$.
- $y_1 w' x_2 = (u_1)^{k_1} v (u_2)^{k_2}$ for some $k_1, k_2 \in \mathbb{N}$.

Recall that by hypothesis, we have a word $z \in A^*$ such that $vz \sim_{\mathcal{C}} u_1$ and $zv \sim_{\mathcal{C}} u_2$. We define $w \in A^*$ as follows,

$$w = x_1(u_1)^p((u_1)^{k_1} v (u_2)^{k_2} z)^p y_1 w' x_2 (u_2)^p y_2$$

We show that $w \in L_iKL_j$ and that there are two decompositions witnessing this membership. Thus, this proves that L_iKL_j is not unambiguous which is a contradiction.

Since $vz \sim_{\mathcal{C}} u_1$, $zv \sim_{\mathcal{C}} u_2$ and $\sim_{\mathcal{C}}$ is a congruence, it follows that,

$$(u_1)^{k_1} v (u_2)^{k_2} z \sim_{\mathcal{C}} (vz)^{k_1} v (zv)^{k_2} z = (vz)^{k_1+k_2+1} \sim_{\mathcal{C}} (u_1)^{k_1+k_2+1}$$

Therefore, by definition of p, Fact 15 yields that,

$$x_1(u_1)^p((u_1)^{k_1} v (u_2)^{k_2} z)^p y_1 \sim_{\mathcal{C}} x_1(u_1)^p y_1 = w_i$$

Finally, since $w_i \in L_i$ which is a language of $\mathcal{C}$, it follows by definition of $\sim_{\mathcal{C}}$ that,

$$x_1(u_1)^p((u_1)^{k_1} v (u_2)^{k_2} z)^p y_1 \in L_i$$

Moreover, we have, $w' \in K$ by definition and $x_2(u_2)^p y_2 = w_j \in L_j$. Therefore,

$$w = x_1(u_1)^p((u_1)^{k_1} v (u_2)^{k_2} z)^p y_1 w' x_2 (u_2)^p y_2 \in L_iKL_j$$

It remains to find a second decomposition of w witnessing this membership.

Recall that we have $y_1 w' x_2 = (u_1)^{k_1} v (u_2)^{k_2}$. Thus, it is simple to verify from the definition of w that,

$$\begin{aligned}
w &= x_1(u_1)^p((u_1)^{k_1} v (u_2)^{k_2} z)^p y_1 w' x_2 (u_2)^p y_2 \\
w &= x_1(u_1)^p((u_1)^{k_1} v (u_2)^{k_2} z)^p (u_1)^{k_1} v (u_2)^{k_2} (u_2)^p y_2 \\
w &= x_1(u_1)^p (u_1)^{k_1} v (u_2)^{k_2} (z(u_1)^{k_1} v (u_2)^{k_2})^p (u_2)^p y_2 \\
w &= x_1(u_1)^p y_1 w' x_2 (z(u_1)^{k_1} v (u_2)^{k_2})^p (u_2)^p y_2
\end{aligned}$$

Using similar arguments to those we applied above, one may then verify that,

$$x_2(z(u_1)^{k_1} v (u_2)^{k_2})^p (u_2)^p y_2 \in L_j$$

Moreover, $x_1(u_1)^p y_1 = w_i \in L_i$ and $w' \in K$. Thus, we get a second decomposition witnessing the membership of w in L_iKL_J which concludes the proof. ◀

C Proof of Theorem 4

In this appendix, we complete the proof of Theorem 4. It remained to prove the directions (1) ⇒ (4) and (4) ⇒ (5) as well as two lemmas in our proof for the direction (5) ⇒ (2). We start by presenting a few results about $\mathcal{C}$-pairs and saturated $\mathcal{C}$-pairs that we shall need for our arguments.

C.1 Preliminary results on $\mathcal{C}$-pairs

Let us briefly recall the definitions of $\mathcal{C}$-pairs and saturated $\mathcal{C}$-pairs. Let $\mathcal{C}$ be some class of languages. Consider a finite monoid M and a *surjective* morphism $\alpha : A^* \to M$. Given a pair $(s,t) \in M \times M$,

- (s,t) is a *$\mathcal{C}$-pair* (for α) when **no** language of $\mathcal{C}$ can separate $\alpha^{-1}(s)$ from $\alpha^{-1}(t)$.
- (s,t) is a *saturated $\mathcal{C}$-pair* (for α) when **no** language of $\mathcal{C}$ **recognized by α** can separate $\alpha^{-1}(s)$ from $\alpha^{-1}(t)$.

We start with a lemma presenting an important property of $\mathcal{C}$-pairs. This relation is reflexive and compatible with multiplication when $\mathcal{C}$ is a quotienting lattice of regular languages (it is also symmetric when $\mathcal{C}$ is a Boolean algebra, but we do not use this property).

▶ **Lemma 17.** *Let $\mathcal{C}$ be a quotienting lattice of regular languages, M be a finite monoid and $\alpha : A^* \to M$ be a surjective morphism. Then, the $\mathcal{C}$-pair relation is reflexive and compatible with multiplication*

Proof. Clearly, the $\mathcal{C}$-pair relation is reflexive: given $s \in M$, the language $\alpha^{-1}(s)$ is nonempty since α is reflexive, whence it cannot be separated from $\alpha^{-1}(s)$. Therefore, (s,s) is a $\mathcal{C}$-pair.

We now prove that the relation is a congruence: for any two $\mathcal{C}$-pairs (s_1,t_1) and (s_2,t_2), (s_1s_2, t_1t_2) is a $\mathcal{C}$-pair as well. We show the contrapositive. Assume that (s_1s_2, t_1t_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 $\alpha^{-1}(s_1s_2) \subseteq K$ and $K \cap \alpha^{-1}(t_1t_2) = \emptyset$. We define,

$$H_1 = \bigcap_{w \in \alpha^{-1}(s_2)} Kw^{-1}$$

By definition, $H_1 \in \mathcal{C}$ since $\mathcal{C}$ is a quotienting lattice and contains only regular languages (thus K has finitely many right quotients by Myhill-Nerode theorem). Additionally, since $\alpha^{-1}(s_1s_2) \subseteq K$, one may verify from the definition that $\alpha^{-1}(s_1) \subseteq H_1$. There are now two cases.

- If $\alpha^{-1}(t_1) \cap H_1 = \emptyset$ then $H_1 \in \mathcal{C}$ separates $\alpha^{-1}(s_1)$ from $\alpha^{-1}(t_1)$ and we are done: (s_1,t_1) is not a $\mathcal{C}$-pair.
- Otherwise, there exists a word $u_1 \in \alpha^{-1}(t_1) \cap H_1 \neq \emptyset$. Let $H_2 = (u_1)^{-1}K \in \mathcal{C}$. We claim that H_2 separates $\alpha^{-1}(s_2)$ from $\alpha^{-1}(t_2)$ which concludes the proof: (s_2,t_2) is not a $\mathcal{C}$-pair. Indeed, given $w \in \alpha^{-1}(s_2)$, we have $u_1 \in H_1 \subseteq Kw^{-1}$ by definition which means that $u_1w \in K$ and therefore that $w \in H_2 = (u_1)^{-1}K$. Therefore, $\alpha^{-1}(s_2) \subseteq H_2$. Now, assume by contradiction that there exists $v \in \alpha^{-1}(t_2) \cap H_2$. Since $H_2 = (u_1)^{-1}K$, it follows that $u_1v \in K$. Finally, since $\alpha(u_1) = t_1$ and $\alpha(v) = t_2$, it follows that $u_1v \in \alpha^{-1}(t_1t_2)$. Therefore, we would have $u_1v \in K \cap \alpha^{-1}(t_1t_2)$, a contradiction since this language is empty by hypothesis.

◀

We now turn to the properties of *saturated* $\mathcal{C}$-pairs. The key result is that when $\mathcal{C}$ is a Boolean algebra, the saturated $\mathcal{C}$-pair relation is an equivalence on M (contrary to the $\mathcal{C}$-pair relation, which is not transitive in general). Moreover, the equivalence classes correspond exactly to the languages recognized by α which belong to $\mathcal{C}$.

▶ Remark. When $\mathcal{C}$ is only a lattice, the saturated $\mathcal{C}$-pair relation is only a preorder and the languages recognized by α which belong to $\mathcal{C}$ correspond to its upper sets. We do not consider this case.

▶ **Lemma 18.** *Let $\mathcal{C}$ be a Boolean algebra, let M be a finite monoid and let $\alpha : A^* \to M$ be a surjective morphism. Then, the saturated $\mathcal{C}$-pair relation is an equivalence relation. Moreover, for any $F \subseteq M$, the two following properties are equivalent:*

1. *$\alpha^{-1}(F) \in \mathcal{C}$.*
2. *F is a union of equivalence classes for the saturated $\mathcal{C}$-pair relation.*

Proof. We first show that that the saturated $\mathcal{C}$-pair relation is an equivalence. Clearly, it is reflexive: given $s \in M$, $\alpha^{-1}(s)$ (which is nonempty since α is surjective) cannot be separated from $\alpha^{-1}(s)$, thus (s, s) is a saturated $\mathcal{C}$-pair.

We turn to transitivity. Consider $r, s, t \in M$ such that (r, s) and (s, t) are saturated $\mathcal{C}$-pairs. We show that (r, t) is a saturated $\mathcal{C}$-pair as well. That is, we must show that no language of $\mathcal{C}$ recognized by α separates $\alpha^{-1}(r)$ from $\alpha^{-1}(t)$. Consider $L \in \mathcal{C}$ recognized by α such that $\alpha^{-1}(r) \subseteq L$. We have to show that $\alpha^{-1}(t) \cap L \neq \emptyset$. Since (r, s) is a saturated $\mathcal{C}$-pair, L cannot separate $\alpha^{-1}(r)$ from $\alpha^{-1}(s)$. Thus, $\alpha^{-1}(s) \cap L \neq \emptyset$. Moreover, since L is recognized by α, this implies that $\alpha^{-1}(s) \subseteq L$. Finally, since (s, t) is a saturated $\mathcal{C}$-pair, L cannot separate $\alpha^{-1}(s)$ from $\alpha^{-1}(t)$. Thus, $\alpha^{-1}(t) \cap L \neq \emptyset$ and we are finished.

Finally, we prove that the saturated $\mathcal{C}$-pair relation is symmetric. We use the fact that $\mathcal{C}$ is closed under complement. Consider a saturated $\mathcal{C}$-pair $(s, t) \in M^2$. We show that (t, s) is a saturated $\mathcal{C}$-pair as well. By contradiction assume that we have $H \in \mathcal{C}(A)$ recognized by α separating $\alpha^{-1}(t)$ from $\alpha^{-1}(s)$. It follows that $A^* \setminus H$ separates $\alpha^{-1}(s)$ from $\alpha^{-1}(t)$. Clearly, $A^* \setminus H$ is still recognized by α and since $\mathcal{C}$ is closed under complement, $A^* \setminus H \in \mathcal{C}$. Thus, this contradicts the hypothesis that (s, t) is a saturated $\mathcal{C}$-pair. This concludes the proof that the saturated $\mathcal{C}$-pair relation is an equivalence.

It remains to prove the equivalence between the two items in the lemma. Consider $F \subseteq M$. We start with the direction (1) $\Rightarrow$ (2). Assume that $\alpha^{-1}(F) \in \mathcal{C}$, we show that F is a union of equivalence classes for the saturated $\mathcal{C}$-pair relation. Consider $s \in F$ and $t \in M$ such that (s, t) is a saturated $\mathcal{C}$-pair, we have to 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 done.

We turn to the direction (2) $\Rightarrow$ (1). Assume that F is a union of equivalence classes for the saturated $\mathcal{C}$-pair relation. We show that $\alpha^{-1}(F) \in \mathcal{C}$. Consider $s \in F$ and $t \notin F$. By the hypothesis on F and since $s \in F$, the whole class of s for the saturated $\mathcal{C}$-pair relation is included in F. Therefore, (s, t) is **not** a saturated $\mathcal{C}$-pair, so that there exists a language $L_{s,t}$ in $\mathcal{C}$ recognized by α that separates $\alpha^{-1}(s)$ from $\alpha^{-1}(t)$, that is, such that $L_{s,t} \in \mathcal{C}$, $\alpha^{-1}(s) \subseteq L_{s,t}$ and $\alpha^{-1}(t) \cap L_{s,t} = \emptyset$. One may then verify that,

$$\alpha^{-1}(F) = \bigcup_{s \in F} \bigcap_{t \notin F} L_{s,t}.$$

Since $\mathcal{C}$ is a closed under union and intersection, it follows that $\alpha^{-1}(F) \in \mathcal{C}$. This concludes the proof. ◀

We may now connect our two relations with the following lemma. The saturated $\mathcal{C}$-pairs are the transitive closure of the $\mathcal{C}$-pairs.

▶ **Lemma 19.** *Consider a Boolean algebra $\mathcal{C}$, a finite monoid M and a surjective morphism $\alpha : A^* \to M$. 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. The direction (2) $\Rightarrow$ (1) is immediate since any $\mathcal{C}$-pair is also a saturated $\mathcal{C}$-pair and we showed in Lemma 18 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 $F \subseteq M$ be 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 18, this will imply that $t \in F$ as well since (s,t) is a saturated $\mathcal{C}$-pair. Thus, Item (2) holds. We now show that $\alpha^{-1}(F) \in \mathcal{C}$. 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 $\alpha^{-1}(v)$ and since $\mathcal{C}$ is closed under intersection, it suffices to define,

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

We now 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 closed under union. ◀

Finally, a corollary of these results is that when $\mathcal{C}$ is a quotienting Boolean algebra of regular languages the saturated $\mathcal{C}$-pair relation is a congruence.

▶ **Lemma 20.** *Let* $\mathcal{C}$ *be a quotienting Boolean algebra of regular languages,* M *a finite monoid and* $\alpha : A^* \to M$ *a surjective morphism. Then, the saturated* $\mathcal{C}$*-pair relation is a congruence.*

Proof. By Lemma 18, we already know that the saturated $\mathcal{C}$-pair relation is an equivalence. Moreover, it is immediate from Lemma 19 that it is compatible with multiplication since we already know that this is the case for the $\mathcal{C}$-pair relation by Lemma 17. ◀

C.2 Direction (1) $\Rightarrow$ (4) in Theorem 4

Recall that a quotienting Boolean algebra of regular languages $\mathcal{C}$ is fixed. Consider a regular language $L \in \mathit{UPol}(\mathcal{C})$ and let $\alpha : A^* \to M$ be its syntactic morphism. Given any $\mathcal{C}$-pair $(s,t) \in M \times M$, we have to show that $s^{\omega+1} = s^\omega t s^\omega$. We first prove the following simple fact.

▶ **Fact 21.** *There exists a finite quotienting Boolean algebra* $\mathcal{D} \subseteq \mathcal{C}$ *such that* $L \in \mathit{UPol}(\mathcal{D})$.

Proof. Since $L \in \mathit{UPol}(\mathcal{C})$, it is built from finitely many languages in $\mathcal{C}$ using disjoint unions and unambiguous marked concatenations. We let $\mathcal{F} \subseteq \mathcal{C}$ as the finite class containing all basic languages in $\mathcal{C}$ used in the construction. Moreover, we let $\mathcal{D}$ as the smallest quotienting Boolean algebra containing $\mathcal{F}$. Clearly $\mathcal{D} \subseteq \mathcal{C}$ since $\mathcal{C}$ is a quotienting Boolean algebra itself. Moreover, $L \in \mathit{UPol}(\mathcal{D})$ since $\mathcal{D}$ contains all languages in $\mathcal{C}$ required to build L by

definition. It remains to show that $\mathcal{D}$ remains finite. By definition, the languages in $\mathcal{D}$ are built from those in $\mathcal{F}$ by applying Boolean operations and quotients. Therefore, since quotients commute with Boolean operations, any language in $\mathcal{D}$ is a Boolean combination of quotients of languages in $\mathcal{F}$. Finally, any regular language has finitely many quotients by Myhill-Nerode theorem. Thus, since $\mathcal{F}$ was finite, this is the case for $\mathcal{D}$ as well. ◀

We work with the canonical equivalence $\sim_{\mathcal{D}}$ over A^* associated to the finite quotienting Boolean algebra $\mathcal{D}$. Since (s,t) is a $\mathcal{C}$-pair, we know that $\alpha^{-1}(s)$ is not $\mathcal{C}$-separable from $\alpha^{-1}(t)$. Therefore, since $\mathcal{D} \subseteq \mathcal{C}$, it follows that $\alpha^{-1}(s)$ is not $\mathcal{D}$-separable from $\alpha^{-1}(t)$. By Lemma 12 the unions of $\sim_{\mathcal{D}}$ classes belong to $\mathcal{D}$. Therefore, some $\sim_{\mathcal{D}}$-class intersects both $\alpha^{-1}(s)$ and $\alpha^{-1}(t)$ (otherwise, the union of all $\sim_{\mathcal{D}}$-classes intersecting $\alpha^{-1}(s)$ would be a separator in $\mathcal{D}$). Hence, we have $u \in \alpha^{-1}(s)$ and $v \in \alpha^{-1}(t)$ such that $u \sim_{\mathcal{D}} v$. Hence, we may apply Corollary 14 which yields a natural number $h \in \mathbb{N}$ such that for any $x, y \in A^*$,

$$xu^{h\omega}vu^{h\omega}y \in L \quad \text{if and only if} \quad xu^{h\omega+1}y \in L$$

This exactly says that $u^{h\omega}vu^{h\omega}$ and $u^{h\omega+1}$ are equivalent for the syntactic congruence $\equiv_L$ of L. By definition of the syntactic morphism, it then follows that,

$$s^{\omega+1} = \alpha(u^{h\omega+1}) = \alpha(u^{h\omega}vu^{h\omega}) = s^{\omega}ts^{\omega}$$

This concludes the proof for this direction.

C.3 Direction (4) ⇒ (5) in Theorem 4

Recall that a quotienting Boolean algebra $\mathcal{C}$ is fixed. Consider a regular language L and let $\alpha : A^* \to M$ be its syntactic morphism. Assume that Equation (4) in Theorem 4 holds: for any $\mathcal{C}$-pair $(s,t) \in M^2$, we have $s^{\omega+1} = s^{\omega}ts^{\omega}$. We have to show that Equation (5) holds as well: for any **saturated** $\mathcal{C}$-pair $(s,t) \in M^2$, we have $s^{\omega+1} = s^{\omega}ts^{\omega}$

Consider a saturated $\mathcal{C}$-pair $(s,t) \in M^2$. We show that $s^{\omega+1} = s^{\omega}ts^{\omega}$. By Lemma 19, 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 $\mathcal{C}$-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 by hypothesis (i.e. Equation (4) holds) that $s^{\omega+1} = s^{\omega}r_1 s^{\omega}$ since (s, r_1) is a $\mathcal{C}$-pair. We now assume that $k > 1$. Using induction, we get that,

$$s^{\omega+1} = s^{\omega}r_{k-1}s^{\omega}$$

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 $\mathcal{C}$-pair, It is immediate from Lemma 17 that, $(s^{\omega}r_{k-1}s^{\omega}, s^{\omega}r_k s^{\omega})$ is a $\mathcal{C}$-pair as well. Thus, since Equation (4) holds, we get that,

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

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

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

This concludes the proof.

C.4 Missing proofs for the direction $(5) \Rightarrow (2)$ in Theorem 4

Recall that a quotienting Boolean algebra $\mathcal{C}$ is fixed. For this direction, it only remained to prove Lemma 6 and Lemma 10. We start with the former.

C.4.1 Proof of Lemma 6

We have a surjective morphism $\alpha : A^* \to M$ in hand and we need to show that there exists a finite monoid N and a surjective morphism $\beta : M \to N$ which satisfies the following properties:

- For any $s, t \in M$, (s, t) is a saturated $\mathcal{C}$-pair if and only if $\beta(s) = \beta(t)$.
- Any language recognized by the composition $\gamma = \beta \circ \alpha : A^* \to N$ belongs to $\mathcal{C}$.

By Lemmas 18 and 20, we now that the saturated $\mathcal{C}$-pair relation is a congruence on M. Therefore, we may define N as the monoid obtained by quotienting M by this congruence and $\beta : M \to N$ as the corresponding morphism (which associates it equivalence class to any element $s \in M$). The first item is then immediate by definition.

For the second item, it follows by definition that any language L recognized by the composition $\beta \circ \alpha : A^* \to N$ is of the form $L = \alpha^{-1}(F)$ where F is a union of equivalence classes for the saturated $\mathcal{C}$-pair relation. Thus, it follows from Lemma 18 that $L \in \mathcal{C}$.

C.4.2 Proof of Lemma 10

Let us first recall the situation. We have two finite monoids M and N together with two surjective morphisms $\alpha : A^* \to M$ and $\beta : M \to N$. Moreover, γ denotes the composition $\beta \circ \alpha : A^* \to N$. We have $r_1, r_2, s \in M$ and any $x \in N$ and we are considering the following language:

$$L_s^x[r_1, r_2] = \{w \in \gamma^{-1}(x) \mid r_1 \cdot \alpha(w) \cdot r_2 = s\}$$

Additionally, we are working under the hypothesis that $x <_{\mathcal{J}} 1_N$. In this situation, we defined T as the set of all triples $(y, a, z) \in N \times A \times N$ satisfying the three following conditions: $x = y \cdot \gamma(a) \cdot z$, $x <_{\mathcal{J}} y$ and $x \mathcal{J} y \cdot \gamma(a)$. Lemma 10 states that $L_s^x[r_1, r_2]$ is equal to the following disjoint union:

$$L_s^x[r_1, r_2] = \biguplus_{(y,a,z) \in T} \left(\biguplus_{t \in \beta^{-1}(y)} \alpha^{-1}(t) \cdot a \cdot L_s^z[r_1 t \alpha(a), r_2] \right)$$

We first prove the above equality holds and then show that the union is disjoint. We start with the right to left inclusion. Given a triple $(y, a, z) \in T$, $t \in \beta^{-1}(y)$ and $w \in \alpha^{-1}(t) \cdot a \cdot L_s^z[r_1 t \alpha(a), r_2]$, we have to show that $w \in L_s^x[r_1, r_2]$. By definition, $w = uav$ with $\alpha(u) = t$ and $v \in L_s^z[r_1 t \alpha(a), r_2]$. The condition on v implies that $\gamma(v) = z$ and $r_1 t \alpha(a) \alpha(v) r_2 = s$. Since $\alpha(u) = t$, we get $r_1 \alpha(w) r_2 = r_1 \alpha(uav) r_2 = s$. Moreover, $\gamma(w) = \gamma(uav) = \beta(t) \cdot \gamma(a) \cdot z = y \cdot \gamma(a) \cdot z = x$ by definition of T. We conclude that $w \in L_s^x[r_1, r_2]$.

We turn to the converse inclusion. Consider $w \in L_s^x[r_1, r_2]$. We have to exhibit $(y, a, z) \in T$ and $t \in \beta^{-1}(y)$ such that $w \in \alpha^{-1}(t) \cdot a \cdot L_s^z[r_1 t \alpha(a), r_2]$. By hypothesis, we have $x <_{\mathcal{J}} 1_N$. Thus, since $\gamma(w) = x$, we know that $w \neq \varepsilon$. Hence, w has a smallest non-empty prefix $u' \in A^+$ such that $\gamma(u') \mathcal{J} x$. We let $u, v \in A^*$ and $a \in A$ such that $u' = ua$ and $w = uav$. Finally, we let $t = \alpha(u)$, $y = \beta(t) = \gamma(u)$ and $z = \gamma(v)$. Clearly, $t \in \beta^{-1}(y)$. Therefore, it remains to show that $(y, a, z) \in T$ and $w \in \alpha^{-1}(t) \cdot a \cdot L_s^z[r_1 t \alpha(a), r_2]$. We start with $(y, a, z) \in T$.

There are three conditions to check. First, it is clear that $y \cdot \gamma(a) \cdot z = \gamma(uav) = \gamma(w) = x$. Moreover, ua is by definition the smallest non-empty prefix of w such that $\gamma(ua) \mathcal{J} x$. Thus, since $y = \gamma(u)$, it is immediate that $x \mathcal{J} y \cdot \gamma(a)$. Finally, it also follows that when u is non-empty, we have $x <_{\mathcal{J}} \gamma(u) = y$ and if $u = \varepsilon$, then $y = \gamma(u) = 1_N$ and we have $x <_{\mathcal{J}} 1_N$ by hypothesis. Finally, we show that $w \in \alpha^{-1}(t) \cdot a \cdot L_s^z[r_1 t\alpha(a), r_2]$. Since $w = uav$ and $t = \alpha(u)$ by definition, it suffices to show that $v \in L_s^z[r_1 t\alpha(a), r_2]$. We have $v \in \gamma^{-1}(z)$ by definition of z. Moreover, since $w \in L_s^x[r_1, r_2]$. We have $r_1\alpha(w)r_2 = s$. As $\alpha(u) = t$ by definition, this yields $r_1 t\alpha(a)\alpha(v)r_2 = s$. Altogether, we obtain that $v \in L_s^z[r_1 t\alpha(a), r_2]$.

It remains to prove that the union is disjoint. Consider $(y, a, z), (y', a', z') \in T$, $t \in \beta^{-1}(y)$ and $t' \in \beta^{-1}(y')$. Moreover, assume that we have a word $w \in A^*$ such that,

$$w \in \left(\alpha^{-1}(t) \cdot a \cdot L_s^z[r_1 t\alpha(a), r_2]\right) \cap \left(\alpha^{-1}(t') \cdot a' \cdot L_s^{z'}[r_1 t'\alpha(a'), r_2]\right)$$

We show that $(y, a, z) = (y', a', z')$ and $t = t'$. Since $w \in \alpha^{-1}(t) \cdot a \cdot L_s^z[r_1 t\alpha(a), r_2]$, it admits a decomposition $w = uav$ with $\alpha(u) = t$ and $v \in L_s^z[r_1 t\alpha(a), r_2]$. Moreover, since $(y, a, z) \in T$, we have $x <_{\mathcal{J}} y = \gamma(u)$ and $x \mathcal{J} y \cdot \gamma(a) = \gamma(ua)$. Thus, ua is the smallest non-empty prefix of w such that $x \mathcal{J} \gamma(ua)$. Symmetrically, since $w \in \alpha^{-1}(t') \cdot a' \cdot L_s^{z'}[r_1 t'\alpha(a'), r_2]$, it admits a decomposition $w = u'a'v'$ with $\alpha(u') = t'$ and $v' \in L_s^{z'}[r_1 t'\alpha(a'), r_2]$. Using the same argument as above, we obtain that $u'a'$ is the smallest non-empty prefix of w such that $x \mathcal{J} \gamma(u'a')$. Thus, we get that $ua = u'a'$ which entails that $u = u'$, $a = a'$ and $v = v'$. This implies as desired that $(y, a, z) = (y', a', z')$ and $t = t'$, finishing the proof.

D Proof of Lemma 13

In this appendix, we prove Lemma 13. Recall that a finite quotienting Boolean algebra $\mathcal{C}$ is fixed. Let us recall the statement of Lemma 13.

▶ **Lemma 13.** *Let $\mathcal{D}$ be a Boolean algebra and assume that there exists an algorithm that takes as input a nice $\mathcal{C}$-compatible rating map $\rho : 2^{A^*} \to R$ and outputs $\mathcal{I}_{\mathcal{D}}[\rho]$. Then, $\mathcal{D}$-separation is decidable.*

We assume that there exists an algorithm which takes as input a nice $\mathcal{C}$-compatible rating map $\rho : 2^{A^*} \to R$ and outputs $\mathcal{I}_{\mathcal{D}}[\rho]$. We describe an algorithm for $\mathcal{D}$-separation. This procedures involves two steps:

1. First, we explain how to compute a $\mathcal{C}$-compatible rating map ρ from two input regular languages L_1 and L_2.
2. Then, we show that $\mathcal{I}_{\mathcal{D}}[\rho]$ (which we may compute by hypothesis) is enough information to decide whether L_1 is $\mathcal{D}$-separable from L_1 and L_2.

When put together, these two steps prove that $\mathcal{D}$-separation is decidable. We fix the two regular languages L_1 and L_2 for the proof. Since they are regular, one may compute two finite monoids M_1, M_2 and two morphisms $\alpha_1 : A^* \to M_1$ and $\alpha_2 : A^* \to M_2$ recognizing L_1 and L_2 respectively.

Construction of the rating map ρ. Recall that since $\mathcal{C}$ is a finite quotienting Boolean algebra, the quotient set $A^*/{\sim_{\mathcal{C}}}$ is a finite monoid and the map $w \mapsto [w]_{\mathcal{C}}$ a morphism. We define M as the following monoid equipped with the component-wise multiplication:

$$M = M_1 \times M_2 \times (A^*/{\sim_{\mathcal{C}}})$$

Furthermore, we define $\alpha : A^* \to M$ as the following morphism,

$$\begin{array}{rccl} \alpha: & A^* & \to & M \\ & w & \mapsto & (\alpha_1(w), \alpha_2(w), [w]_{\mathcal{C}}) \end{array}$$

By definition α recognizes both L_1 and L_2. We now consider the rating map, $\rho : 2^{A^*} \to 2^M$ defined by,

$$\begin{array}{rccl} \rho: & 2^{A^*} & \to & 2^M \\ & K & \mapsto & \{\alpha(w) \mid w \in K\} \end{array}$$

It is straightforward to verify that ρ is a nice rating map (this is true for any morphism $\alpha : A^* \to M$). Clearly, one may compute it from L_1 and L_2. Moreover, by construction of α, ρ is $\mathcal{C}$-compatible.

▶ **Fact 14.** *The rating map ρ is $\mathcal{C}$-compatible.*

Proof. Consider two $\sim_{\mathcal{C}}$-classes C and D. We have to show that when there exists $S \in 2^M \setminus \{\emptyset\}$ such that $S \subseteq \rho(C)$ and $S \subseteq \rho(D)$, then $C = D$. Since $S \neq \emptyset$, we have $s \in S$. By definition $s \in M$ is a triple $s = (s_1, s_2, E)$ where E is a $\sim_{\mathcal{C}}$-class. We show that $E = C = D$. By symmetry, it suffices to prove that $E = C$. Since $S \subseteq \rho(C)$, we have $(s_1, s_2, E) \in \rho(C)$ and by definition of ρ, this yields $w \in C$ such that $\alpha(w) = (s_1, s_2, E)$. Thus, by definition of α, we have $[w]_{\mathcal{C}} = E$. Since $w \in C$, this yields $C = E$. ◀

Connection with $\mathcal{D}$-separation. In view of Fact 14, we may apply our hypothesis algorithm to compute $\mathcal{I}_{\mathcal{D}}[\rho]$. We now show in the next lemma that this information suffices to decide whether L_1 is $\mathcal{D}$-separable from L_2 which concludes the proof of Lemma 13.

▶ **Lemma 15.** *The two following properties are equivalent:*

1. *L_1 is $\mathcal{D}$-separable from L_2.*
2. *For any $s_1 \in \alpha(L_1)$ and $s_2 \in \alpha(L_2)$, we have $\{s_1, s_2\} \notin \mathcal{I}_{\mathcal{D}}[\rho]$.*

The remainder of this appendix is devoted to proving Lemma 15. Assume first that L_1 is $\mathcal{D}$-separable from L_2. Given $s_1 \in \alpha(L_1)$ and $s_2 \in \alpha(L_2)$, we have to show that $\{s_1, s_2\} \notin \mathcal{I}_{\mathcal{D}}[\rho]$. By hypothesis, we have $K \in \mathcal{D}$ such that $L_1 \subseteq K$ and $L_2 \cap K = \emptyset$. Clearly, $\mathbf{K} = \{K, A^* \setminus K\}$ is a universal $\mathcal{D}$-cover (note that $A^* \setminus K \in \mathcal{D}$ because $\mathcal{D}$ is a Boolean algebra). Hence, by definition of $\mathcal{I}_{\mathcal{D}}[\rho]$, we have,

$$\mathcal{I}_{\mathcal{D}}[\rho] \subseteq \mathcal{I}[\rho](\mathbf{K})$$

Hence, it suffices to show that $\{s_1, s_2\} \notin \mathcal{I}[\rho](\mathbf{K})$. By contradiction assume that $\{s_1, s_2\} \in \mathcal{I}[\rho](\mathbf{K})$. By definition, this means that either $\{s_1, s_2\} \subseteq \rho(K)$ or $\{s_1, s_2\} \subseteq \rho(A^* \setminus K)$.

- In the former case, we get $\{s_2\} \subseteq \rho(K)$ which means that there exists $w \in K$ such that $\alpha(w) = s_2$. Since $s_2 \in \alpha(L_2)$ and L_2 is recognized by α, this implies $w \in L_2$. This is a contradiction as we know that $L_2 \cap K = \emptyset$ by hypothesis.
- In the latter case, we get $\{s_1\} \subseteq \rho(A^* \setminus K)$ which means that there exists $w \notin\in K$ such that $\alpha(w) = s_1$. Since $s_1 \in \alpha(L_1)$ and L_1 is recognized by α, this implies $w \in L_1$. This is a contradiction as we know that $L_1 \subseteq K$ by hypothesis.

We turn to the converse direction. Assume that for any $s_1 \in \alpha(L_1)$ and $s_2 \in \alpha(L_2)$, we have $\{s_1, s_2\} \notin \mathcal{I}_{\mathcal{D}}[\rho]$. We show that L_1 is $\mathcal{D}$-separable from L_2. Let $\mathbf{K}$ be an optimal universal $\mathcal{D}$-cover for ρ. In other words, $\mathcal{I}[\rho](\mathbf{K}) = \mathcal{I}_{\mathcal{D}}[\rho]$. Moreover, we define,

$$H = \bigcup_{\{K \in \mathbf{K} \mid K \cap L_1 \neq \emptyset\}} K$$

Clearly $H \in \mathcal{D}$ since it is a union of languages in $\mathbf{K}$ which is a $\mathcal{D}$-cover. We show that H separates L_1 from L_2, finishing the proof.

It is immediate that $L_1 \subseteq H$ by definition since $\mathbf{K}$ is a universal $\mathcal{D}$-cover which means that any word in A^* (and in particular any word in L_1) belongs to some language $K \in \mathbf{K}$. Thus, we may concentrate on proving that $L_2 \cap H = \emptyset$. By contradiction Assume that there exists $w_2 \in L_2 \cap H$. By definition of H, this yields some $K \in \mathbf{K}$ such that $K \cap L_1 \neq \emptyset$ and $w_2 \in K$. Therefore, K contains $w_2 \in L_2$, and $w_1 \in L_1$. By definition of ρ, this implies that $\{\alpha(w_1), \alpha(w_2)\} \subseteq \rho(K)$. Moreover, by definition of $\mathcal{I}[\rho](\mathbf{K})$, this yields $\{\alpha(w_1), \alpha(w_2)\} \in \mathcal{I}[\rho](\mathbf{K}) = \mathcal{I}_{\mathcal{D}}[\rho]$. This is a contradiction. Indeed, since $w_1 \in L_1$ and $w_2 \in L_2$, we have $\alpha(w_1) \in \alpha(L_1)$ and $\alpha(w_2) \in \alpha(L_2)$. In this case, our hypothesis states that $\{\alpha(w_1), \alpha(w_2)\} \notin \mathcal{I}_{\mathcal{D}}[\rho]$.

E Proof of Theorem 14

This appendix is devoted to proving Theorem 14. Recall that $\mathcal{C}$ is a fixed finite quotienting Boolean algebra. Given a rating map $\rho : 2^{A^*} \to R$ and a subset S of R, we say that S is $UPol(\mathcal{C})$-saturated (for ρ) if it contains the set $\mathcal{I}_{triv}[\rho] = \{r \in R \mid r \leq \rho(w) \text{ for some } w \in A^*\}$ and is closed under the following operations:

1. *Downset*: for any $s \in S$, if $r \in R$ satisfies $r \leq s$, then we have $r \in S$.
2. *Multiplication*: For any $s, t \in S$, we have $st \in S$.
3. *$UPol(\mathcal{C})$-closure*: Given two $\sim_{\mathcal{C}}$-classes C, D and $s, t \in S$ such that $s \leq \rho(C \bullet D)$ and $t \leq \rho(D \bullet C)$, we have $s^\omega \cdot \rho(C) \cdot t^\omega \in S$

The statement of Theorem 14 is as follows.

▶ **Theorem 14.** *Let $\rho : 2^{A^*} \to R$ be a $\mathcal{C}$-compatible rating map. Then, $\mathcal{I}_{UPol(\mathcal{C})}[\rho]$ is the smallest $UPol(\mathcal{C})$-saturated subset of R (with respect to inclusion).*

Therefore, we fix a $\mathcal{C}$-compatible rating map ρ. The proof is organized around two separate arguments which correspond respectively to soundness and completeness in the least fixpoint procedure computing $\mathcal{I}_{UPol(\mathcal{C})}[\rho] \subseteq R$:

- First, we show that $\mathcal{I}_{UPol(\mathcal{C})}[\rho]$ is $UPol(\mathcal{C})$-saturated. This is the soundness part of the proof: the least fixpoint procedure only computes elements of $\mathcal{I}_{UPol(\mathcal{C})}[\rho]$.
- Then, we show that $\mathcal{I}_{UPol(\mathcal{C})}[\rho]$ is smaller than any $UPol(\mathcal{C})$-saturated subset. This is the completeness part of the proof: our least fixpoint procedure computes all elements in $\mathcal{I}_{UPol(\mathcal{C})}[\rho]$. This direction is of particular interest as it describes a generic procedure for building optimal $UPol(\mathcal{C})$-covers.

E.1 Soundness

We prove the soundness part of Theorem 14. In other words, our objective is to show that $\mathcal{I}_{UPol(\mathcal{C})}[\rho]$ is $UPol(\mathcal{C})$-saturated. We show that it contains $\mathcal{I}_{triv}[\rho]$ and is closed under downset, multiplication and $UPol(\mathcal{C})$-closure.

Since $UPol(\mathcal{C})$ is a quotienting Boolean algebra of regular languages (see Theorem 3), we already know that $\mathcal{I}_{UPol(\mathcal{C})}[\rho]$ contains $\mathcal{I}_{triv}[\rho]$ and is closed under downset and multiplication. Indeed, it was shown in [17] that this is a generic property of quotienting Boolean algebras of regular languages. Hence, we may concentrate on proving that $\mathcal{I}_{UPol(\mathcal{C})}[\rho]$ satisfies $UPol(\mathcal{C})$-closure. Consider two $\sim_{\mathcal{C}}$-classes C, D and $s, t \in \mathcal{I}_{UPol(\mathcal{C})}[\rho]$ such that $s \leq \rho(C \bullet D)$ and $t \leq \rho(D \bullet C)$. We have to show that,

$$s^\omega \cdot \rho(C) \cdot t^\omega \in \mathcal{I}_{UPol(\mathcal{C})}[\rho]$$

Let us first observe that we may assume without loss of generality that $s \neq 0_R$ and $t \neq 0_R$. Indeed, otherwise, $s = 0_R$ or $t = 0_R$ and it is immediate that $s^\omega \cdot \rho(C) \cdot t^\omega = 0_R \in \mathcal{I}_{UPol(\mathcal{C})}[\rho]$. Hence, we assume from now on that $s, t \neq 0_R$.

By definition, proving that $s^\omega \cdot \rho(C) \cdot t^\omega \in \mathcal{I}_{UPol(\mathcal{C})}[\rho]$ amounts to showing that for any universal $UPol(\mathcal{C})$-cover $\mathbf{K}$, we have $s^\omega \cdot \rho(C) \cdot t^\omega \in \mathcal{I}[\rho](\mathbf{K})$. Therefore, we fix an arbitrary universal $UPol(\mathcal{C})$-cover $\mathbf{K}$ for the proof. By definition, we have to exhibit $K \in \mathbf{K}$ such that $s^\omega \cdot \rho(C) \cdot t^\omega \leq \rho(K)$. We start with a few definitions that we need to describe K.

Since $\mathcal{C}$ and $\mathbf{K}$ are finite and $UPol(\mathcal{C})$ is a quotienting Boolean algebra of regular languages, we have the following fact.

▶ **Fact 15.** *There exists a finite quotienting Boolean algebra $\mathcal{D}$ such that $\mathcal{C} \subseteq \mathcal{D} \subseteq UPol(\mathcal{C})$ and any $K \in \mathbf{K}$ belongs to $\mathcal{D}$.*

Proof. We simply define $\mathcal{D}$ as the smallest quotienting Boolean algebra which contains all languages in $\mathbf{K}$ and $\mathcal{C}$. Clearly, $\mathcal{C} \subseteq \mathcal{D} \subseteq UPol(\mathcal{C})$ since $UPol(\mathcal{C})$ is a quotienting Boolean algebra by Theorem 3 and all $K \in \mathbf{K} \cup \mathcal{C}$ belong to $UPol(\mathcal{C})$. Moreover, any $K \in \mathbf{K}$ belongs to $\mathcal{D}$ by definition. Therefore, it suffices to show that $\mathcal{D}$ is finite. It is simple to verify that Boolean operations commute with quotients. Thus, we get by definition that any language in $\mathcal{D}$ is a Boolean combination of languages of the form $w^{-1}K$ or Kw^{-1} where $K \in \mathbf{K} \cup \mathcal{C}$ and $w \in A^*$. Thus, since any language in $\mathbf{K} \cup \mathcal{C}$ has finitely many quotients (by Myhill-Nerode theorem, this holds for any regular language), we obtain that $\mathcal{D}$ is a finite quotienting Boolean algebra. ◀

Recall that we write $\sim_{\mathcal{D}}$ for the canonical equivalence on A^* associated to $\mathcal{D}$ (it compares words belonging to the same languages in $\mathcal{D}$). Since $\mathcal{D}$ is closed under quotients we known that $\sim_{\mathcal{D}}$ is a congruence for word concatenation by Lemma 12. Moreover, by definition of $\sim_{\mathcal{D}}$ and Proposition 13, we have the following lemma.

▶ **Lemma 16.** *There exists $h \in \mathbb{N}$ such that for any $\ell \geq h$ and $u_1, u_2, v, v', z \in A^*$ satisfying $v \sim_{\mathcal{C}} v'$, $vz \sim_{\mathcal{C}} v'z \sim_{\mathcal{C}} u_1$ and $zv \sim_{\mathcal{C}} zv' \sim_{\mathcal{C}} u_2$, we have,*

$$(u_1)^\ell v (u_2)^\ell \sim_{\mathcal{D}} (u_1)^\ell v' (u_2)^\ell$$

Proof. Proposition 13 yields that for any language $L \in \mathcal{D} \subseteq UPol(\mathcal{C})$, there exists $h_L \in \mathbb{N}$ such that for any $\ell \geq h$ and $u_1, u_2, v, v', z \in A^*$ satisfying $v \sim_{\mathcal{C}} v'$, $vz \sim_{\mathcal{C}} v'z \sim_{\mathcal{C}} u_1$ and $zv \sim_{\mathcal{C}} zv' \sim_{\mathcal{C}} u_2$, we have,

$$(u_1)^\ell v (u_2)^\ell \in L \quad \text{if and only if} \quad (u_1)^\ell v' (u_2)^\ell \in L$$

We simply choose h as the maximum of all numbers h_L for $L \in \mathcal{D}$ (recall that $\mathcal{D}$ is finite). It then follows, that for any $\ell \geq h$ and $u_1, u_2, v, v', z \in A^*$ satisfying $v \sim_{\mathcal{C}} v'$, $vz \sim_{\mathcal{C}} v'z \sim_{\mathcal{C}} u_1$ and $zv \sim_{\mathcal{C}} zv' \sim_{\mathcal{C}} u_2$, we have,

$$(u_1)^\ell v (u_2)^\ell \in L \quad \text{if and only if} \quad (u_1)^\ell v' (u_2)^\ell \in L \quad \text{for any } L \in \mathcal{D}$$

By definition, this exactly says that $(u_1)^\ell v (u_2)^\ell \sim_{\mathcal{D}} (u_1)^\ell v' (u_2)^\ell$. ◀

We may now come back to the main argument. Recall that we have two $\sim_{\mathcal{C}}$-classes C, D and $s, t \in \mathcal{I}_{UPol(\mathcal{C})}[\rho]$ such that $s \leq \rho(C \bullet D)$, $t \leq \rho(D \bullet C)$, $s \neq 0_R$ and $t \neq 0_R$. Moreover, we have a universal $UPol(\mathcal{C})$-cover $\mathbf{K}$ and we want to find $K \in \mathbf{K}$ such that $s^\omega \cdot \rho(C) \cdot t^\omega \leq \rho(K)$. We construct K.

Let $\mathbf{H}$ be the partition of A^* into $\sim_{\mathcal{D}}$-classes. Clearly, $\mathbf{H}$ is a universal $\mathcal{D}$-cover and therefore a universal $UPol(\mathcal{C})$-cover as well since we have $\mathcal{D} \subseteq LDet(\mathcal{C})$ by definition. Therefore, since $s, t \in \mathcal{I}_{UPol(\mathcal{C})}[\rho]$, we have $s, t \in \mathcal{I}[\rho](\mathbf{H})$ and by definition, we get two $\sim_{\mathcal{D}}$-classes $H_s, H_t \in \mathbf{H}$ such that, $s \leq \rho(H_s)$ and $t \leq \rho(H_t)$. Consider the natural number $h \in \mathbb{N}$ given by Lemma 16. We define,

$$G = (H_s)^{h\omega} \cdot C \cdot (H_t)^{h\omega}$$

Note that since $H_s, H_t \in \mathbf{H}$ are $\sim_{\mathcal{D}}$-classes and C is a $\sim_{\mathcal{C}}$-class, G is non-empty. The argument is now based on the following lemma which we use to build the desired language $K \in \mathbf{K}$ such that $s^\omega \cdot \rho(C) \cdot t^\omega \leq \rho(K)$.

▶ **Lemma 17.** *The language G is included in a $\sim_{\mathcal{D}}$-class.*

Before we prove Lemma 17, let us use it to conclude the argument. Since G is non-empty and $\mathbf{K}$ is a universal cover, there exists some $K \in \mathbf{K}$ such that $K \cap G \neq \emptyset$. Let $w \in K \cap G$. By Lemma 17 any word in G is $\sim_{\mathcal{D}}$-equivalent to w. Therefore, since $K \in \mathcal{D}$ by definition of $\mathcal{D}$ and $w \in K$, it follows that $G \subseteq K$. Consequently, we have,

$$(\rho(H_s))^\omega \cdot \rho(C) \cdot (\rho(H_t))^\omega = \rho(G) \leq \rho(K)$$

Since $s \leq \rho(H_s)$ and $t \leq \rho(H_t)$ by definition of H_s and H_t, we get as desired that $s^\omega \cdot \rho(C) \cdot t^\omega \leq \rho(K)$ which concludes the proof.

It remains to prove Lemma 17. We show that G is included in a $\sim_{\mathcal{D}}$-class. We start by proving a property of the languages H_s and H_t. This is where we use the hypothesis that ρ is $\mathcal{C}$-compatible (in fact, this is the only place in the whole proof of Theorem 14 where we use this hypothesis).

▶ **Fact 18.** *We have $H_s \subseteq C \bullet D$ and $H_t \subseteq D \bullet C$.*

Proof. We present a proof for H_s, the argument is identical for H_t. By definition, H_s is included in a $\sim_{\mathcal{D}}$-class. Moreover, we know that $\mathcal{D}$ includes $\mathcal{C}$ (by definition in Fact 15). Thus, H_s is included in some $\sim_{\mathcal{C}}$-class U. It now suffices to show that $U = C \bullet D$.

Since $H_s \subseteq U$, we have $\rho(H_s) \leq \rho(U)$. Moreover, $s \leq \rho(H_s)$ by definition of H_s. Thus, $s \leq \rho(U)$. Finally, we also have $s \neq 0_R$ and $s \leq \rho(C \bullet D)$ by definition of s. Since ρ is $\mathcal{C}$-compatible, this implies that $U = C \bullet D$. ◀

We now consider some words $u_1 \in H_s$ and $u_2 \in H_t$. It is immediate from Fact 18 that $u_1 \in C \bullet D$ and $u_2 \in D \bullet C$. Moreover, let $v \in C$. We prove that any word in G is $\sim_{\mathcal{D}}$-equivalent to the word $w = (u_1)^{h\omega} v (u_2)^{h\omega}$. This concludes the proof: we obtain as desired that G is included in the $\sim_{\mathcal{D}}$-class of w.

Consider a word $x \in K$. By definition of G, x is of the form $x = u_1' v' u_2'$ with $u_1' \in (H_s)^{h\omega}$, $u_2' \in (H_t)^{h\omega}$ and $v' \in C$. We prove independently that the following two properties hold:

$$x = u_1' v' u_2' \sim_{\mathcal{D}} (u_1)^{h\omega} v' (u_2)^{h\omega} \quad \text{and} \quad (u_1)^{h\omega} v' (u_2)^{h\omega} \sim_{\mathcal{D}} (u_1)^{h\omega} v (u_2)^{h\omega} = w$$

By transitivity, it will then be immediate that $x \sim_{\mathcal{D}} w$, concluding the soundness proof.

We start with the left equivalence. Recall that by hypothesis, H_s is included in a $\sim_{\mathcal{D}}$-class. Therefore, all words in H_s are $\sim_{\mathcal{D}}$-equivalent to $u_1 \in H_s$. Hence, since $u_1' \in (H_s)^{h\omega}$ and $\sim_{\mathcal{D}}$ is a congruence, we obtain that $u_1' \sim_{\mathcal{D}} (u_1)^{h\omega}$. Symmetrically, one may verify that $u_2' \sim_{\mathcal{D}} (u_2)^{h\omega}$. Finally, using again the fact that $\sim_{\mathcal{D}}$ is a congruence again, we obtain the desired property:

$$x = u_1'v'u_2' \sim_{\mathcal{D}} (u_1)^{h\omega}v'(u_2)^{h\omega}$$

Finally, the right equivalence is immediate from Lemma 16. Indeed, we have $h \leq h\omega$. Moreover, let $z \in C$. Since $v, v' \in C$, $u_1 \in C \bullet D$ and $u_2 \in D \bullet C$, we have $v \sim_{\mathcal{C}} v'$, $vz \sim_{\mathcal{C}} v'z \sim_{\mathcal{C}} u_1$ and $zv \sim_{\mathcal{C}} zv' \sim_{\mathcal{C}} u_2$. Therefore, Lemma 16 yields that,

$$(u_1)^{h\omega}v'(u_2)^{h\omega} \sim_{\mathcal{D}} (u_1)^{h\omega}v(u_2)^{h\omega}$$

This concludes the soundness proof.

E.2 Completeness

e turn to completeness in Theorem 14. Recall that we have a $\mathcal{C}$-compatible multiplicative rating map $\rho : 2^{A^*} \to R$. Consider a subset $S \subseteq R$ which is $UPol(\mathcal{C})$-saturated. Our objective is to prove that $\mathcal{I}_{UPol(\mathcal{C})}[\rho] \subseteq S$. We rely on the usual approach and construct a universal $UPol(\mathcal{C})$-cover $\mathbf{K}$ such that $\mathcal{I}[\rho](\mathbf{K}) \subseteq S$. By definition of $\mathcal{I}_{UPol(\mathcal{C})}[\rho]$, it will then follow that,

$$\mathcal{I}_{UPol(\mathcal{C})}[\rho] \subseteq \mathcal{I}[\rho](\mathbf{K}) \subseteq S$$

▶ Remark. Since we already showed that $\mathcal{I}_{UPol(\mathcal{C})}[\rho]$ itself is $UPol(\mathcal{C})$-saturated, one may apply the construction when $S = \mathcal{I}_{UPol(\mathcal{C})}[\rho]$. We get a universal $UPol(\mathcal{C})$-cover $\mathbf{K}$ such that $\mathcal{I}_{UPol(\mathcal{C})}[\rho] \subseteq \mathcal{I}[\rho](\mathbf{K}) \subseteq \mathcal{I}_{UPol(\mathcal{C})}[\rho]$. In other words, this builds an optimal $UPol(\mathcal{C})$-cover.

▶ Remark. We do not use the fact that ρ is $\mathcal{C}$-compatible. This was only needed for the soundness proof.

As expected, we build our universal $UPol(\mathcal{C})$-cover $\mathbf{K}$ by induction. To present this construction, we need a new notion called *$\mathcal{C}$-guarded pair* which we present now.

$\mathcal{C}$-guarded pairs. A $\mathcal{C}$-guarded pair is an element $(s, D) \in S \times (A^*/\sim_{\mathcal{C}})$ (i.e. $s \in S$ and D is a $\sim_{\mathcal{C}}$-class) such that $s \leq \rho(D)$. We shall write $P \subseteq S \times (A^*/\sim_{\mathcal{C}})$ for the set of all $\mathcal{C}$-guarded pairs. We have the following useful fact which states that P is a monoid for the component-wise multiplication.

▶ **Fact 19.** *The set P of $\mathcal{C}$-guarded pairs is a monoid for the component-wise multiplication. More precisely, given two $\mathcal{C}$-guarded pairs $(s, D), (t, E) \in P$, the pair $(st, D \bullet E)$ is $\mathcal{C}$-guarded as well. Moreover, the pair $(1_R, [\varepsilon]_{\mathcal{C}})$ belongs to P and is a neutral element for this multiplication.*

Proof. We first show closure under multiplication. Let $(s, D), (t, E) \in P$, we show that $(st, D \bullet E) \in P$. Since $s, t \in S$ and S is closed under multiplication (it is $UPol(\mathcal{C})$-saturated) we have $st \in S$. Thus, it suffices to verify that $st \leq \rho(D \bullet E)$. We know that $s \leq \rho(D)$ and $t \leq \rho(E)$ by definition. Thus, $st \leq \rho(DE)$. Finally, $DE \subseteq D \bullet E$ which yields $st \leq \rho(D \bullet E)$. It remains to show that $(1_R, [\varepsilon]_{\mathcal{C}}) \in P$. By definition, $1_R \leq \rho(\varepsilon)$ which means that $1_R \in \mathcal{I}_{triv}[\rho]$. Therefore, $1_R \in S$ since S is $UPol(\mathcal{C})$-saturated. Moreover, since $\varepsilon \in [\varepsilon]_{\mathcal{C}}$, we have $1_R \leq \rho(\varepsilon) \leq \rho([\varepsilon]_{\mathcal{C}})$. Altogether, we get that $(1_R, [\varepsilon]_{\mathcal{C}}) \in P$. ◀

In view of Fact 19, we shall write "$\cdot$" for the multiplication of P. given $(s, D), (t, E) \in P$, we write $(s, D) \cdot (t, E)$ for $(st, D \bullet E) \in P$.

Main argument. We may now start building our universal $UPol(\mathcal{C})$-cover $\mathbf{K}$ such that $\mathcal{I}[\rho](\mathbf{K}) \subseteq S$. The construction is based on the following proposition which we prove by induction. Recall that a cover $\mathbf{H}$ of some language L is *tight* when $H \subseteq L$ for all $H \in \mathbf{H}$ (in other words, $L = \bigcup_{H \in \mathbf{H}} H$).

▶ **Proposition 20.** *Let C be a $\sim_{\mathcal{C}}$-class and $(s_1, D_1), (s_2, D_2) \in P$. There exists a tight $UPol(\mathcal{C})$-cover $\mathbf{H}$ of C such that,*

$$\text{For all } H \in \mathbf{H}, \quad s_1 \cdot \rho(H) \cdot s_2 \in S \tag{1}$$

Before we prove Proposition 20, we use it to build our universal $UPol(\mathcal{C})$-cover $\mathbf{K}$ and finish the completeness proof. We know from Fact 19 that $(1_R, [\varepsilon]_{\mathcal{C}}) \in P$. Hence, given any $\sim_{\mathcal{C}}$-class C, we may apply Proposition 20 in the special case when $(s_1, D_1) = (s_2, D_2) = (1_R, [\varepsilon]_{\mathcal{C}})$. This yields a $UPol(\mathcal{C})$-cover $\mathbf{K}_C$ of C such that $\rho(K) \in S$ for all $K \in \mathbf{K}_C$. We now define $\mathbf{K}$ as the union,

$$\mathbf{K} = \bigcup_{C \in (A^*/\sim_{\mathcal{C}})} \mathbf{K}_C$$

Since the union of all $\sim_{\mathcal{C}}$-classes is A^*, $\mathbf{K}$ is clearly a universal $UPol(\mathcal{C})$-cover. We have to verify that $\mathcal{I}[\rho](\mathbf{K}) \subseteq S$. Consider $r \in \mathcal{I}[\rho](\mathbf{K})$. By definition, there exists $K \in \mathbf{K}$ such that $r \leq \rho(K)$. By definition of $\mathbf{K}$, we have $\rho(K) \in S$. Moreover, S is closed under downset (it is $UPol(\mathcal{C})$-saturated) which yields $r \in S$. This concludes the completeness proof for Theorem 14.

It remains to prove Proposition 20. We devote the remainder of the section to this proof. Let $C \in A^*/\sim_{\mathcal{C}}$ be a $\sim_{\mathcal{C}}$-class and $(s_1, D_1), (s_2, D_2) \in P$ be two $\mathcal{C}$-guarded pairs. We have to build a tight $UPol(\mathcal{C})$-cover $\mathbf{H}$ of C satisfying (1). The construction is an induction on three parameters that we define now.

Since the set $A^*/\sim_{\mathcal{C}}$ of $\sim_{\mathcal{C}}$-classes is a monoid, we may consider its Green relations. We define the *rank* of some $\sim_{\mathcal{C}}$-class F as the number of $\sim_{\mathcal{C}}$-classes U which are larger than F for $\leqslant_{\mathcal{J}}$ (i.e. $F \leqslant_{\mathcal{J}} U$). Our first parameter is the rank of $D_1 \bullet C \bullet D_2$.

Similarly, since P is a monoid by Fact 19, we may consider its Green relations as well. Given $p \in P$, we define the *right index* of p as number of $\mathcal{C}$-guarded pairs $q \in P$ which are smaller than p for $\leqslant_{\mathcal{R}}$ (i.e. $q \leqslant_{\mathcal{R}} p$). Symmetrically, the *left index* of p is the number of $\mathcal{C}$-guarded pairs $q \in P$ which are smaller than p for $\leqslant_{\mathcal{L}}$ (i.e. $q \leqslant_{\mathcal{L}} p$). We build our tight $UPol(\mathcal{C})$-cover $\mathbf{H}$ of C satisfying (1) by induction on the three following parameters listed by importance:

1. The rank of $D_1 \bullet C \bullet D_2$.
2. The right index of (s_1, D_1).
3. The left index of (s_2, D_2).

▶ Remark. It is crucial that the rank of $D_1 \bullet C \bullet D_2$ is a more important induction parameter than the other two. Indeed, it may happen that the second and third parameters increase when we use induction on the rank of $D_1 \bullet C \bullet D_2$. On the other hand, the order between the other two parameters is arbitrary.

We may now begin the proof and construct $\mathbf{H}$. We distinguish three cases depending on the following properties of C, (s_1, D_1) and (s_2, D_2).

- We say that C is *smooth* when $C \mathcal{J} D_1 \bullet C \bullet D_2$.

- We say that (s_1, D_1) is *right stable* when there exists $(t, E) \in P$ such that $E \mathrel{\Re} C$ and $(s_1, D_1) \cdot (t, E) \mathrel{\Re} (s_1, D_1)$.
- We say that (s_2, D_2) is *left stable* when there exists $(t, E) \in P$ such that $E \mathrel{\mathcal{L}} C$ and $(t, E) \cdot (s_2, D_2) \mathrel{\mathcal{L}} (s_2, D_2)$.

In the base case, we assume that the three above properties hold and use them to build $\mathbf{H}$ directly. Otherwise, we consider two inductive cases. First, we consider the case when C is not smooth which we handle by induction on our first parameter (the rank of $D_1 \bullet C \bullet D_2$). Finally, we treat the case when either (s_1, D_1) is not right stable or (s_2, D_2) is not left stable using induction on all parameters.

E.2.1 Base case in Proposition 20

We assume that C is smooth, (s_1, D_1) is right stable and (s_2, D_2) is left stable. In this case, we simply define $\mathbf{H} = \{C\}$. Clearly, this is a tight cover of C. Moreover, since C is a $\sim_{\mathcal{C}}$-class, we have $C \in \mathcal{C} \subseteq \mathit{UPol}(\mathcal{C})$. Thus, $\mathbf{H}$ is a tight $\mathit{UPol}(\mathcal{C})$-cover of C as desired. It now remains to prove that (1) holds. Since $\mathbf{H}$ contains only the language C, this amounts to showing that,

$$s_1 \cdot \rho(C) \cdot s_2 \in S$$

Let us first use our hypothesis to define objects that we shall need to prove this property.

- Since the pair (s_1, D_1) is right stable, there exists $(t_1, E_1) \in P$ such that $E_1 \mathrel{\Re} C$ and $(s_1, D_1) \cdot (t_1, E_1) \mathrel{\Re} (s_1, D_1)$. In particular, the second property yields another pair $(x_1, F_1) \in P$ such that,

$$(s_1, D_1) = (s_1, D_1) \cdot (t_1 x_1, E_1 \bullet F_1)$$

- Since the pair (s_2, D_2) is left stable, there exists $(t_2, E_2) \in P$ such that $E_2 \mathrel{\mathcal{L}} C$ and $(t_2, E_2) \cdot (s_2, D_2) \mathrel{\mathcal{L}} (s_2, D_2)$. In particular, the second property yields another pair $(x_2, F_2) \in P$ such that,

$$(s_2, D_2) = (x_2 t_2, F_2 \bullet E_2) \cdot (s_2, D_2)$$

The argument is now based on the following lemma (which is proved using the hypothesis that C is smooth).

▶ **Lemma 21.** *There exists some $\sim_{\mathcal{C}}$-class U such that $E_1 \bullet F_1 = C \bullet U$ and $F_2 \bullet E_2 = U \bullet C$.*

Before we prove the lemma, let us use it to finish the proof of the base case. We show that $s_1 \cdot \rho(C) \cdot s_2 \in S$. By definition, we have $(t_1 x_1, E_1 \bullet F_1) \in P$ and $(x_2 t_2, F_2 \bullet E_2) \in P$. In view of Lemma 21 this yields that, $(t_1 x_1, C \bullet U) \in P$ and $(x_2 t_2, U \bullet C) \in P$. Therefore, by definition of P, we have $t_1 x_1, x_2 t_2 \in S$, $t_1 x_1 \leq \rho(C \bullet U)$ and $x_2 t_2 \leq \rho(U \bullet C)$. This is exactly the situation in which $\mathit{UPol}(\mathcal{C})$-closure may be applied. This yields,

$$(t_1 x_1)^\omega \cdot \rho(C) \cdot (x_2 t_2)^\omega \in S$$

Moreover, by definition, we have $s_1 = s_1 t_1 x_1$ and $s_2 = x_2 t_2 s_2$. Therefore, we obtain $s_1 = s_1 (t_1 x_1)^\omega$ and $s_2 = (x_2 t_2)^\omega s_2$. Hence,

$$s_1 \cdot \rho(C) \cdot s_2 = s_1 (t_1 x_1)^\omega \cdot \rho(C) \cdot (x_2 t_2)^\omega s_2$$

Since we already have $(t_1 x_1)^\omega \cdot \rho(C) \cdot (x_2 t_2)^\omega \in S$ and $s_1, s_2 \in S$ (they are both part of a $\mathcal{C}$-guarded pair), closure under multiplication then yields as desired that $s_1 \cdot \rho(C) \cdot s_2 \in S$. This concludes our argument for the base case. It remains to prove Lemma 21.

Proof of Lemma 21. Since C is smooth, we know that $D_1 \bullet C \bullet D_2 \mathrel{\mathcal{J}} C$. In particular, this implies that $D_1 \bullet C \mathrel{\mathcal{J}} C$ and $C \bullet D_2 \mathrel{\mathcal{J}} C$. Since it is clear that $D_1 \bullet C \leqslant_{\mathcal{L}} C$ and $C \bullet D_2 \leqslant_{\mathcal{R}} C$, Lemma 1 then yields $D_1 \bullet C \mathrel{\mathcal{L}} C$ and $C \bullet D_2 \mathrel{\mathcal{R}} C$. Moreover, by definition, we have $E_1 \mathrel{\mathcal{R}} C$ and $E_2 \mathrel{\mathcal{L}} C$. Altogether, we get $D_1 \bullet C \mathrel{\mathcal{L}} E_2$ and $C \bullet D_2 \mathrel{\mathcal{R}} E_1$. This yields two $\sim_{\mathcal{C}}$-classes G_1, G_2 such that,

$$E_1 = C \bullet D_2 \bullet G_2 \quad \text{and} \quad E_2 = G_1 \bullet D_1 \bullet C$$

Moreover, we also have the following equalities by hypothesis: $D_1 = D_1 \bullet E_1 \bullet F_1$ and $D_2 = F_2 \bullet E_2 \bullet D_2$. We may now replace E_1 and E_2 in these expressions which yields,

$$\begin{aligned} D_1 &= D_1 \bullet C \bullet D_2 \bullet G_2 \bullet F_1 \\ D_2 &= F_2 \bullet G_1 \bullet D_1 \bullet C \bullet D_2 \end{aligned}$$

We may now replace D_1 and D_2 in our expressions for E_1 and E_2 which yields,

$$\begin{aligned} E_1 \bullet F_1 &= C \bullet F_2 \bullet G_1 \bullet D_1 \bullet C \bullet D_2 \bullet G_2 \bullet F_1 \\ F_2 \bullet E_2 &= F_2 \bullet G_1 \bullet D_1 \bullet C \bullet D_2 \bullet G_2 \bullet F_1 \bullet C \end{aligned}$$

Finally, we define U as the following $\sim_{\mathcal{C}}$-class: $U = F_2 \bullet G_1 \bullet D_1 \bullet C \bullet D_2 \bullet G_2 \bullet F_1$. This simplifies our expressions for $E_1 \bullet F_1$ and $F_2 \bullet E_2$ as follows,

$$\begin{aligned} E_1 \bullet F_1 &= C \bullet U \\ F_2 \bullet E_2 &= U \bullet C \end{aligned}$$

This concludes the proof. ◀

E.2.2 First inductive case in Proposition 20

In this first inductive case, we assume that C is not smooth. We handle it using induction on our first parameter: the rank of $D_1 \bullet C \bullet D_2$.

It is clear that we have $D_1 \bullet C \bullet D_2 \leqslant_{\mathcal{J}} C$. Moreover, since C is not smooth, we know that C and $D_1 \bullet C \bullet D_2$ are not $\mathcal{J}$-equivalent which means the the inequality is strict: $D_1 \bullet C \bullet D_2 <_{\mathcal{J}} C = [\varepsilon]_{\mathcal{C}} \bullet C \bullet [\varepsilon]_{\mathcal{C}}$. Altogether, it follows that the rank of $[\varepsilon]_{\mathcal{C}} \bullet C \bullet [\varepsilon]_{\mathcal{C}}$ is strictly smaller than the one of $D_1 \bullet C \bullet D_2$. Therefore since $(1_R, [\varepsilon]_{\mathcal{C}}) \in P$, we may apply induction on our first parameter to get a tight $UPol(\mathcal{C})$-cover $\mathbf{H}$ of C such that $\rho(H) \in S$ for all $H \in \mathbf{H}$. Finally, since $s_1, s_2 \in S$ (they are both part of a $\mathcal{C}$-guarded pair), it then follows by closure under multiplication that $s_1 \cdot \rho(H) \cdot s_2 \in S$ for all $H \in \mathbf{H}$. Hence, $\mathbf{H}$ is a tight $UPol(\mathcal{C})$-cover of C satisfying (1) and we are finished.

E.2.3 Second inductive case in Proposition 20

We now assume that either (s_1, D_1) is not right stable or (s_2, D_2) is not left stable. Thus we have two sub-cases. Since they are symmetrical we detail the one when (s_2, D_2) is not left stable and leave the other to the reader.

▶ Remark. Similarly to what happened when proving the algebraic characterization of $UPol(\mathcal{C})$ (i.e. Theorem 4), since we focus on the case when (s_2, D_2) is not left stable, we only apply induction on our first and third parameters (the rank of $D_1 \bullet C \bullet D_2$ and the left index of (s_2, D_2)). Moreover, we build the languages in our cover $\mathbf{H}$ from those obtained by induction using *right deterministic marked concatenations* only. As expected, handling the dual case ((s_1, D_1) is not right stable) requires using induction on our second parameter (the right index of (s_1, D_1)) and left deterministic marked concatenations.

Observe that our hypothesis yields the following weaker property which will be useful in the construction.

▶ **Fact 22.** *We have* $C <_{\mathcal{J}} [\varepsilon]_{\mathcal{C}}$.

Proof. Clearly, we have $C \leqslant_{\mathcal{J}} [\varepsilon]_{\mathcal{C}}$ (the neutral element is maximal for any Green relation). By contradiction, assume that $C \not<_{\mathcal{J}} [\varepsilon]_{\mathcal{C}}$ which means that $C \mathcal{J} [\varepsilon]_{\mathcal{C}}$. Since $C \leqslant_{\mathcal{L}} [\varepsilon]_{\mathcal{C}}$, this implies $C \mathcal{L} [\varepsilon]_{\mathcal{C}}$ by Lemma 1. It now follows that $(1_R, [\varepsilon]_{\mathcal{C}}) \in P$ satisfies $C \mathcal{L} [\varepsilon]_{\mathcal{C}}$ and $(1_R, [\varepsilon]_{\mathcal{C}}) \cdot (s_2, D_2) = (s_2, D_2) \mathcal{L} (s_2, D_2)$. This means that (s_2, D_2) is left stable which contradicts our hypothesis. ◀

Recall that our objective is to build a tight $UPol(\mathcal{C})$-cover $\mathbf{H}$ of C satisfying (1). We first decompose C as the union of simpler languages which we shall cover independently. We define T as the set of all triples $(U, a, V) \in (A^*/{\sim_{\mathcal{C}}}) \times A \times (A^*/{\sim_{\mathcal{C}}})$ (i.e. U, V are $\sim_{\mathcal{C}}$-classes and a is a letter) such that,

$$U \bullet [a]_{\mathcal{C}} \bullet V = C, \quad C <_{\mathcal{J}} V \quad \text{and} \quad C \mathcal{J} [a]_{\mathcal{C}} \bullet V$$

It turns out that we may decompose C according to the triples in T. We prove this in the following fact.

▶ **Fact 23.** *We have the following equality,*

$$C = \bigcup_{(U,a,V) \in T} UaV$$

Proof. We start with the right to left inclusion. Let $(U, a, V) \in T$ and $w \in UaV$, we show that $w \in C$. We have $w = uav$ with $u \in U$ and $v \in V$. Thus, $[u]_{\mathcal{C}} = U$ and $[v]_{\mathcal{C}} = V$ which yields that $[w]_{\mathcal{C}} = [uav]_{\mathcal{C}} = U \bullet [a]_{\mathcal{C}} \bullet V$. Since $U \bullet [a]_{\mathcal{C}} \bullet V = C$ by definition of T, we get that $[w]_{\mathcal{C}} = C$ which means that $w \in C$.

Conversely, let $w \in C$. We have to find $(U, a, V) \in T$ such that $w \in UaV$. By Fact 22, we know that $w \neq \varepsilon$. Hence, $w \in C$ has a smallest non-empty suffix $v' \in A^+$ such that $[v']_{\mathcal{C}} \mathcal{J} C$. We let $u, v \in A^*$ and $a \in A$ such that $v' = av$ and $w = uav$. Clearly, $w \in [u]_{\mathcal{C}} a \cdot [v]_{\mathcal{C}}$. Thus, it suffices to verify that $([u]_{\mathcal{C}}, a, [v]_{\mathcal{C}}) \in T$. Clearly, we have $[u]_{\mathcal{C}} \bullet [a]_{\mathcal{C}} \bullet [v]_{\mathcal{C}} = [w]_{\mathcal{C}} = C$. Moreover, av is by definition the smallest non-empty suffix of w such that $[av]_{\mathcal{C}} \mathcal{J} C$. Thus, it is immediate that $C \mathcal{J} [a]_{\mathcal{C}} \bullet [v]_{\mathcal{C}}$. Finally, it also follows that when u is non-empty, we have $C <_{\mathcal{J}} [v]_{\mathcal{C}}$ and if $u = \varepsilon$, $C <_{\mathcal{J}} [v]_{\mathcal{C}}$ follows from Fact 22. ◀

In view of Fact 23, it now suffices to build an individual tight $UPol(\mathcal{C})$-cover $\mathbf{H}_{U,a,V}$ of UaV satisfying (1) for each triple $(U, a, V) \in T$. This is what we do in the following lemma using our hypothesis that (s_2, D_2) is not left stable.

▶ **Lemma 24.** *For any* $(U, a, V) \in T$*, there exists a tight* $UPol(\mathcal{C})$*-cover* $\mathbf{H}_{U,a,V}$ *of* UaV *such that* $s_1 \cdot \rho(H) \cdot s_2 \in S$ *for all* $H \in \mathbf{H}_{U,a,V}$.

Before we prove the lemma, we use it to finish the proof of Proposition 20. We simply define,

$$\mathbf{H} = \bigcup_{(U,a,V) \in T} \mathbf{H}_{U,a,V}$$

Let us verify that $\mathbf{H}$ is indeed a tight $UPol(\mathcal{C})$-cover $\mathbf{H}$ of C satisfying (1). It is immediate from Fact 23 that $\mathbf{H}$ is a tight $UPol(\mathcal{C})$-cover of C since each $\mathbf{H}_{U,a,V}$ is a tight $UPol(\mathcal{C})$-cover of HaV by definition in Lemma 24. Finally, given any $H \in \mathbf{H}$ we have $H \in \mathbf{H}_{U,a,V}$ for some $(U, a, V) \in T$. By definition of $\mathbf{H}_{U,a,V}$ in Lemma 24, this yields $s_1 \cdot \rho(H) \cdot s_2 \in S$. Thus, (1) is satisfied which concludes the argument. We finish with the proof of Lemma 24.

Proof of Lemma 24. We fix a triple $(U, a, V) \in T$ for the proof. Our objective is to build a tight $UPol(\mathcal{C})$-cover $\mathbf{H}_{U,a,V}$ of $UaV \subseteq A^*$ such that $s_1 \cdot \rho(H) \cdot s_2 \in S$ for all $H \in \mathbf{H}_{U,a,V}$.

We first use induction on our first parameter (the rank of $D_1 \bullet C \bullet D_2$) to build a $UPol(\mathcal{C})$-cover $\mathbf{K}_V$ of V. Recall that $(1_R, [\varepsilon]_{\mathcal{C}}) \in P$. Clearly, we have $D_1 \bullet C \bullet D_2 \leqslant_{\mathcal{J}} C$. Moreover, by definition of T, we have $C <_{\mathcal{J}} V = [\varepsilon]_{\mathcal{C}} \bullet V \bullet [\varepsilon]_{\mathcal{C}}$. Altogether, this means that $D_1 \bullet C \bullet D_2 <_{\mathcal{J}} [\varepsilon]_{\mathcal{C}} \bullet V \bullet [\varepsilon]_{\mathcal{C}}$ which implies that the rank of $[\varepsilon]_{\mathcal{C}} \bullet V \bullet [\varepsilon]_{\mathcal{C}}$ is strictly smaller than the one of $D_1 \bullet C \bullet D_2$. Therefore, induction on our first parameter yields a tight $UPol(\mathcal{C})$-cover $\mathbf{K}_V$ of V such that,

$$\rho(K) \in S \quad \text{for all } K \in \mathbf{K}_V \tag{2}$$

We now use our hypothesis that (s_2, D_2) is not left stable to build several tight $UPol(\mathcal{C})$-covers of U, one for each $K \in \mathbf{K}_V$. We present the construction in the following fact. This is where we use induction on the left index of (s_2, D_2) (i.e. our third parameter). Moreover, it is also important here that the covers we obtain by induction are tight.

▶ **Fact 25.** *For all $K \in \mathbf{K}_V$, there exists a tight $UPol(\mathcal{C})$-cover $\mathbf{M}_K$ of U such that,*

For all $M \in \mathbf{M}_K$, $\quad s_1 \cdot \rho(M) \cdot \rho(aK) \cdot s_2 \in S$

Proof. First observe that we have $(\rho(aK), [a]_{\mathcal{C}} \bullet V) \in P$. Indeed, $\rho(K) \in S$ by (2) and since $\mathbf{K}_V$ is a tight cover of V, we have $K \subseteq V$ which implies that $\rho(K) \leq \rho(V)$. Therefore, we get $(\rho(K), V) \in P$. Moreover, we have $\rho(a) \in \mathcal{I}_{triv}[\rho]$ by definition which implies that $\rho(a) \in S$ since S is $UPol(\mathcal{C})$-saturated. Additionally, $a \in [a]_{\mathcal{C}}$ which implies that $\rho(a) \leq \rho([a]_{\mathcal{C}})$. Therefore, $(\rho(a), [a]_{\mathcal{C}}) \in P$ and since P is a monoid, we get $(\rho(aK), [a]_{\mathcal{C}} \bullet V) \in P$.

We use on induction our third parameter in Proposition 20 to build $\mathbf{M}_K$. Specifically, we want to apply the proposition to the $\sim_{\mathcal{C}}$-class U and the $\mathcal{C}$-guarded pairs, $(s_1, D_1) \in P$ (which remains unchanged) and $(\rho(aK), [a]_{\mathcal{C}} \bullet V) \cdot (s_2, D_2) \in P$. This will yield the desired tight $UPol(\mathcal{C})$-cover $\mathbf{M}_K$ of V such that,

For all $M \in \mathbf{M}_K$, $\quad s_1 \cdot \rho(M) \cdot \rho(aK) \cdot s_2 \in S$

We now prove that it is possible to apply induction on our third parameter. We first show that our first two parameters have not increased. Then, we show that the third one has decreased which concludes the proof. Since $(U, a, V) \in T$, we know that $U \bullet [a]_{\mathcal{C}} \bullet V = C$ which implies that $D_1 \bullet U \bullet [a]_{\mathcal{C}} \bullet V \bullet D_2 = D_1 \bullet C \bullet D_2$. Therefore, our first induction parameter (the rank of $D_1 \bullet C \bullet D_2$) remains unchanged. Moreover, since we keep using $(s_1, D_1) \in P$, our second parameter (the right index of (s_1, D_1)) remains unchanged as well.

It remains to prove that the left index of $(\rho(aK), [a]_{\mathcal{C}} \bullet V) \cdot (s_2, D_2)$ is strictly smaller than the one of (s_2, D_2). Since $(U, a, V) \in T$, we know that $U \bullet [a]_{\mathcal{C}} \bullet V = C$ (which implies $C \leqslant_{\mathcal{L}} [a]_{\mathcal{C}} \bullet V$) and $C \mathcal{J} [a]_{\mathcal{C}} \bullet V$. By Lemma 1, this yields that $C \mathcal{L} [a]_{\mathcal{C}} \bullet V$. Therefore, since $(\rho(aK), [a]_{\mathcal{C}} \bullet V) \in P$ and (s_2, D_2) is **not** left stable, it is immediate that $(\rho(aK), [a]_{\mathcal{C}} \bullet V) \cdot (s_2, D_2)$ and (s_2, D_2) cannot be $\mathcal{L}$-equivalent. Moreover, since it is clear that we have $(\rho(aK), [a]_{\mathcal{C}} \bullet V) \cdot (s_2, D_2) \leqslant_{\mathcal{L}} (s_2, D_2)$, it follows that this inequality is strict: $(\rho(aK), [a]_{\mathcal{C}} \bullet V) \cdot (s_2, D_2) <_{\mathcal{L}} (s_2, D_2)$. This implies as desired that the left index of $(\rho(aK), [a]_{\mathcal{C}} \bullet V) \cdot (s_2, D_2)$ is strictly smaller than the one of (s_2, D_2) by definition. This concludes the proof. ◀

We may now build $\mathbf{H}_{U,a,V}$ and finish the proof of Lemma 24. For all $K \in \mathbf{K}_V$, we let $\mathbf{M}_K$ as the tight $UPol(\mathcal{C})$-cover of U described in Fact 25. We define $\mathbf{H}_{U,a,V}$ as follows,

$$\mathbf{H}_{U,a,V} = \bigcup_{K \in \mathbf{K}_V} \{MaK \mid M \in \mathbf{M}_K\}$$

Let us verify that $\mathbf{H}_{U,a,V}$ is a tight $UPol(\mathcal{C})$-cover of UaV such that $s_1 \cdot \rho(H) \cdot s_2 \in S$ for all $H \in \mathbf{H}_{U,a,V}$. First observe that $\mathbf{H}_{U,a,V}$ is a tight cover of UaV by definition since $\mathbf{K}_V$ is a tight cover of V and all $\mathbf{M}_K$ for $K \in \mathbf{K}_V$ are tight covers of U.

We now show that any $H \in \mathbf{H}_{U,a,V}$ belongs to $UPol(\mathcal{C})$. We have $H = GaK$ with $K \in \mathbf{K}_V$ and $M \in \mathbf{M}_K$. In particular, we have $K, M \in UPol(\mathcal{C})$ by definition of $\mathbf{K}_V$ and $\mathbf{M}_K$. We prove that MaK is a right deterministic marked concatenation which yields $MaK \in UPol(\mathcal{C})$ as desired. We need to show that $A^*aK \cap K = \emptyset$. Since $\mathbf{K}_V$ is a *tight* cover of V, we have $K \subseteq V$. Thus, it suffices to show that $A^*aV \cap V = \emptyset$. This is because $(U, a, V) \in T$ which implies that $[a]_{\mathcal{C}} \bullet V <_{\mathcal{J}} V$.

Finally, we prove that $s_1 \cdot \rho(H) \cdot s_2 \in S$ for all $H \in \mathbf{H}_{U,a,V}$. Let $H \in \mathbf{H}_{U,a,V}$, we have $H = MaK$ with $K \in \mathbf{K}_V$ and $M \in \mathbf{M}_K$. By definition of $\mathbf{M}_K$, we have $s_1 \cdot \rho(M) \cdot \rho(aK) \cdot s_2 \in S$. This exactly says that $s_1 \cdot \rho(H) \cdot s_2 = s_1 \cdot \rho(MaK) \cdot s_2 \in S$, finishing the proof. ◀