SEPARATING REGULAR LANGUAGES WITH TWO QUANTIFIER ALTERNATIONS

THOMAS PLACE

LaBRI, Bordeaux University, France
e-mail address: tplace@labri.fr

Abstract. We investigate a famous decision problem in automata theory: separation. Given a class of language $\mathcal{C}$, the separation problem for $\mathcal{C}$ takes as input two regular languages and asks whether there exists a third one which belongs to $\mathcal{C}$, includes the first one and is disjoint from the second. Typically, obtaining an algorithm for separation yields a deep understanding of the investigated class $\mathcal{C}$. Therefore, a lot of effort has been devoted to finding algorithms for the most prominent classes.

Here, we are interested in classes within concatenation hierarchies. These hierarchies share a generic construction process: one starts from an initial class called the basis and builds new levels by applying generic operations. The most famous one is the dot-depth hierarchy of Brzozowski and Cohen. It classifies the languages definable in first-order logic. By a theorem of Thomas, it corresponds exactly to the quantifier alternation hierarchy of first-order logic: each level in the dot-depth corresponds to the languages that can be defined with a prescribed number of quantifier blocks. One of the most famous open problems in automata theory is to obtain separation algorithms for all levels in this hierarchy.

Our main theorem is generic: separation is decidable for the level $\frac{3}{2}$ of any concatenation hierarchy whose basis is finite. Moreover, we are able to push this result to the level $\frac{5}{2}$ for the dot-depth. In logical terms, this means that separation is decidable for Σ_3: first-order sentences having at most three quantifier blocks starting with an existential one.

1. Introduction

Context. This paper is part of a research program whose objective is to precisely understand prominent classes of regular languages. Specifically, those defined by descriptive formalisms (such as logic). Naturally, "understanding" a class $\mathcal{C}$ is an informal objective. In the literature, this question is usually approached with a decision problem called *membership*. Given a class $\mathcal{C}$, one wants a procedure deciding whether some input regular language belongs to $\mathcal{C}$. In practice, having such an algorithm in hand yields a deep understanding of the investigated class $\mathcal{C}$: it is an effective description of *all* languages in $\mathcal{C}$.

This approach was originally inspired by theorem of Schützenberger [Sch65] which has two elementary formulations: a language-theoretic one and a logical one. The language-theoretic one gives an algorithm which decides whether an input regular language is star-free,

1998 ACM Subject Classification: F.4.3.

Key words and phrases: Words, regular languages, concatenation hierarchies, first-order logic, quantifier alternation, separation, covering.

LOGICAL METHODS
IN COMPUTER SCIENCE

DOI:10.2168/LMCS-???

i.e., can be expressed with a regular expression using union, complement and concatenation, but not Kleene star. This result was highly influential for several reasons:

- It cemented membership as the "right" question when aiming to "understand" a given class of languages.
- Schützenberger developed a general methodology for tackling membership problems (which he applied to the star-free languages).
- Finally, McNaughton and Papert [MP71] later established that the star-free languages are exactly those which can be defined in first-order logic (FO).

This success motivated researchers to investigate membership for other important classes of languages. This was quite fruitful and the question is now well-understood for several important classes. Prominent examples include the class of piecewise testable languages which was solved by Simon [Sim75] or the two-variables fragment of first-order logic which was solved by Thérien and Wilke [TW98]. However, for some other classes, membership remains wide open, despite a wealth of research work spanning several decades.

Concatenation hierarchies. In the paper, we investigate one of the most famous such open questions: the *dot-depth problem* (see [Pin17] for a survey on the topic). Schützenberger's results motivated Brzozowski and Cohen [BC71] to define an infinite [BK78] classification of all star-free languages: the *dot-depth hierarchy*. Intuitively, one classifies the languages according to the number of alternations between concatenations and complements required to define them. More precisely, the dot-depth is a particular instance of a generic construction process (which was formalized later) named *concatenation hierarchies*. Such a hierarchy has only one parameter: a "level 0 class" called its *basis*. Once it is fixed, one builds new levels by applying two generic operations, polynomial and Boolean closure. There are two kinds of levels: half levels $\frac{1}{2}, \frac{3}{2}, \frac{5}{2} \dots$ and full levels $0, 1, 2, 3 \dots$. For any full level n, the next half and full levels are built as follows:

- Level $n + \frac{1}{2}$ is the *polynomial closure* of level n. Given a class $\mathcal{C}$, its polynomial closure $Pol(\mathcal{C})$ is the smallest class of languages containing $\mathcal{C}$ and closed under union, intersection and marked concatenation $K, L \mapsto KaL$, where a is a letter.
- Level $n+1$ is the *Boolean closure* of level $n+\frac{1}{2}$. Given a class $\mathcal{C}$, its Boolean closure $Bool(\mathcal{C})$ is the smallest class containing $\mathcal{C}$ and closed under union and complement.

Thus, a concatenation hierarchy is fully determined by its basis. In the paper, we are interested in hierarchies with a *finite* basis.

There are essentially two prominent hierarchies of this kind: the dot-depth [BC71] and Straubing-Thérien hierarchies [Str81, Thé81]. This status is partially explained by their connection with quantifier alternation hierarchies. A first-order logic sentence is Σ_n either if its prenex normal form has at most n quantifier blocks and starts with an existential one. Furthermore, a sentence is $\mathcal{B}\Sigma_n$ if it is a finite Boolean combination of Σ_n sentences. It was shown by Thomas [Tho82] that the dot-depth coincides with quantifier alternation: for any $n \in \mathbb{N}$, dot-depth n corresponds to $\mathcal{B}\Sigma_n$ and dot-depth $n + \frac{1}{2}$ to Σ_{n+1}. This result was later lifted to the Straubing-Thérien hierarchy [PP86]: it corresponds to the alternation hierarchy within a variant of first-order logic equipped with a slightly different signature (this does not change the overall expressive power but impacts the languages that one may define at a given level of its alternation hierarchy).

A famous open problem in automata theory is to obtain membership algorithms for all levels in these two hierarchies (and therefore, for the alternation hierarchies of first-order logic as well). However, progress has been slow: until recently, only level $\frac{1}{2}$ [Arf87, PW97], level 1 [Sim75, Kna83] and level $\frac{3}{2}$ [Arf87, PW97, GS00] were solved. See [DGK08] for a survey. Following these results, membership for level 2 remained open for a long time and became famous under the name "dot-depth two problem".

Separation. Recently solutions were found in [PZ14] for the levels 2 and $\frac{5}{2}$ by relying on a new approach. The techniques involve considering a new decision problem stronger than membership: *separation*. Rather than asking whether a single input language belongs to the class $\mathcal{C}$ under investigation, the $\mathcal{C}$-separation problem takes as input *two* regular languages. It asks whether there exists a third language *belonging to* $\mathcal{C}$ which contains the first language and is disjoint from the second. The interest in separation is recent. However, it has quickly replaced membership as the central question. This newly acquired status is explained by two main reasons.

First, separation serves as the key ingredient in all recent membership results (see [PZ15b, PZ17a] for an overview). A simple but crucial example is the membership algorithm of [PZ14] for the level $\frac{5}{2}$ in the Straubing-Thérien hierarchy. This result is based on two ingredients:

(1) A generic reduction from membership for the level $n + \frac{3}{2}$ to separation for the level $n + \frac{1}{2}$ for any full level n in the Straubing-Thérien hierarchy.
(2) A separation algorithm for the level $\frac{3}{2}$.

When combined, these two results immediately yield a membership algorithm for the level $\frac{5}{2}$.

However, there is another deeper reason for working with separation instead of membership. Our primary motivation for considering such decision problems is to thoroughly understand the classes that we investigate. In this respect, while harder, separation is also far more rewarding than membership. Intuitively, this is easily explained. Having a membership algorithm for some class $\mathcal{C}$ in hand only yields benefits for the languages of $\mathcal{C}$: we are able to detect them and to build a description witnessing this membership. Instead, separation algorithms are universal: their benefits apply to ***all*** languages. An insightful perspective is to view separation as an approximation problem: given an input pair (L_1, L_2), we want to over-approximate L_1 by a language in $\mathcal{C}$ and L_2 serves to specify what the acceptable approximations are. Hence, key idea is that looking at separation yields a more robust understanding of the classes.

Separation is known to be decidable for the lower levels in the Straubing-Thérien and dot-depth hierarchies. It was first shown to be decidable for the levels $\frac{1}{2}$, 1, $\frac{3}{2}$ and 2 in the Straubing-Thérien hierarchy [CMM13, PvRZ13, PZ14, PZ17c]. Let us point out that while the results were first formulated independently, it was recently proved in [PZ17c] that the four of them are corollaries of only two generic theorems:

- For any *finite* class $\mathcal{C}$, $Pol(\mathcal{C})$-separation is decidable. The levels $\frac{1}{2}$ and $\frac{3}{2}$ in the Straubing-Thérien hierarchy are of this form (by definition for the former and by a result of [PS85] for the latter).
- For any *finite* class $\mathcal{C}$, $Bool(Pol(\mathcal{C}))$-separation is decidable. The levels 1 and 2 in the Straubing-Thérien hierarchy are of this form (again, this is a result of [PS85] for the latter).

This generic approach was introduced in [PZ17c] and is very important for the paper: our main theorem is based on a similar approach. Finally, it is known that any separation result for some level in the Straubin-Thérien hierarchy may be lifted to the corresponding level in the dot-depth via a generic reduction [PZ15a]. Thus, separation is also decidable for the levels $\frac{1}{2}$, 1, $\frac{3}{2}$ and 2 in the dot-depth.

Contributions. Our most important result in the paper is a separation algorithm for the level $\frac{7}{2}$ in the Straubing-Thérien hierarchy. Note that from logical point of view, this is the level Σ_3 in the quantifier alternation hierarchy of first-order logic: sentences with at most three blocks of quantifiers (i.e. two alternations between quantifiers) and an existential leftmost block. Moreover, we also obtain two important corollaries "for free" from previously known results. First, we are able to lift the result to dot-depth $\frac{7}{2}$ using the aforementioned transfer theorem [PZ15a]. Second, we obtain from the reduction of [PZ14] that *membership* is decidable for the level $\frac{9}{2}$ in the Straubing-Thérien hierarchy (this may lifted to dot-depth as well using a result of Straubing [Str85]).

A crucial point is that this result is **not** our main theorem but only its most important corollary. The actual main result is more general for two distinct reasons. First, we actually consider an even stronger problem that separation: *covering*. This third problem was introduced in [PZ16, PZ17b]. We show that covering is decidable for the level $\frac{7}{2}$ in the Straubing-Thérien hierarchy: the separation result is an immediate corollary. Considering covering has benefits: it is arguably even more rewarding than separation and it comes with an elegant framework [PZ16, PZ17b] designed for tackling it (which we use to formulate all algorithms in the paper). However, we also consider covering out of necessity: no "direct" separation algorithm is known for the level $\frac{7}{2}$.

Moreover, we follow a generic approach similar to the one used in [PZ17c] for $Pol(\mathcal{C})$ and $Bool(Pol(\mathcal{C}))$. Specifically, our main theorem states that covering and separation are decidable for any class of the form $Pol(Bool(Pol(\mathcal{C})))$ with $\mathcal{C}$ a *finite* class (the level $\frac{7}{2}$ in the Straubing-Thérien hierarchy is such a class [PS85]). Being generic, this approach yields algorithms for a whole family of classes. Moreover, we are able to pinpoint the key hypotheses which are critical in order to solve separation and covering for the level $\frac{7}{2}$.

Finally, we shall also reprove the theorem of [PZ17c] for classes of the form $Pol(\mathcal{C})$: when $\mathcal{C}$ is finite, the $Pol(\mathcal{C})$-covering problem is decidable. In fact, we prove a stronger theorem that we need for handling $Pol(Bool(Pol(\mathcal{C})))$.

Organization. Section 2 gives preliminary definitions. We introduce classes of languages and present the decision problems that we consider. In Section 3, we define concatenation hierarchies, state our generic separation theorem and discuss its consequences. The remainder of the paper is devoted to proving this theorem. First, we introduce some mathematical tools that we shall need in Section 4. Sections 5 and 6 are then devoted to presenting the framework that we use for formulating our algorithms. Finally, we present and prove our algorithms for $Pol(\mathcal{C})$- and $Pol(Bool(Pol(\mathcal{C})))$-separation (when $\mathcal{C}$ is finite) in Sections 7 and 8 respectively

This paper is the journal version of [Pla15]. There are several differences between the two versions. In the conference version, the point of view is purely logical: the main theorem is the decidability of Σ_3-separation and concatenation hierarchies are not discussed. In this new version, introducing the language theoretic point of view and concatenation hierarchies allows us to prove a generalized variant of this theorem: $Pol(Bool(Pol(\mathcal{C})))$-separation is decidable for any finite class $\mathcal{C}$. Moreover, while the core ideas remain the same, the proof arguments and their formulation have been significantly modified in order to simplify the presentation and better pinpoint the key ideas.

2. Preliminaries

In this section, we present the objects that we investigate in the paper. We define classes of languages as well as the decision problems that we shall consider. Moreover, we introduce some standard terminology that we shall use.

2.1. Classes of languages.

For the whole paper, we fix an arbitrary finite alphabet A. We shall denote by A^* the set of all words over A, including the empty word ε. We let $A^+ = A^* \setminus \{\varepsilon\}$. If $u, v \in A^*$ are words, we write $u \cdot v$ or uv their concatenation.

A subset of A^* is called a *language*. For the sake of avoiding clutter we often denote the singleton language $\{u\}$ by u. It is standard to extend the concatenation operation to languages: for $K, L \subseteq A^*$, KL denotes 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. All classes that we consider in the paper satisfy robust *closure properties* which we define now.

- A *lattice* is a class of languages $\mathcal{C}$ closed under finite union and intersection, and such that $\emptyset \in \mathcal{C}$ and $A^* \in \mathcal{C}$.
- A *Boolean algebra* is a lattice closed under complement.
- Finally, we say that a class $\mathcal{C}$ is *quotienting* when it is closed under quotients, *i.e.*, when for all $L \in \mathcal{C}$ and all $w \in A^*$, the following two languages both belong to $\mathcal{C}$,

$$\begin{array}{lcl} w^{-1}L & \stackrel{\text{def}}{=} & \{u \in A^* \mid wu \in L\} \text{ and} \\ Lw^{-1} & \stackrel{\text{def}}{=} & \{u \in A^* \mid uw \in L\} \end{array}$$

In the paper, all classes that we consider are at least *quotienting lattices*. Moreover, we only consider classes which are included in the class of *regular languages* (which is something that we shall implicitly assume from now on). These are the languages that can be equivalently defined by monadic second-order logic, finite automata or finite monoids (we come back to this point in Section 5).

2.2. Separation and covering.

Our objective in the paper is to study several specific classes of languages (which we present in Section 3). For this purpose, we shall rely on two decision problems: separation and covering. Both of them are parametrized by an arbitrary class of languages $\mathcal{C}$. Let us start with the definition of separation.

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

INPUT: Two regular languages L_1 and L_2.
OUTPUT: Is L_1 $\mathcal{C}$-separable from L_2 ?

We intend to use separation as a mathematical tool whose purpose is to investigate classes of languages: given a fixed class $\mathcal{C}$, obtaining a $\mathcal{C}$-separation algorithm usually requires a solid understanding of $\mathcal{C}$. In particular, a typical objective when considering separation is to not only get an algorithm that decides it, but also a generic method for computing a separator, when it exists.

Remark 2.1. *$\mathcal{C}$-separation generalizes another classical decision problem: $\mathcal{C}$-membership which asks whether a single regular language L belongs to $\mathcal{C}$. Indeed, it is simple to verify that asking whether $L \in \mathcal{C}$ is equivalent to asking whether L is $\mathcal{C}$-separable from its complement (in this case, the only candidate for being a separator is L itself).*

Covering. Our second problem is more general and was originally defined in [PZ16, PZ17b]. Given a language L, a *cover of* L is a *finite* set of languages $\mathbf{K}$ such that,

$$L \subseteq \bigcup_{K \in \mathbf{K}} K$$

Moreover, given a class $\mathcal{C}$, a $\mathcal{C}$-cover of L is a cover $\mathbf{K}$ of L such that all $K \in \mathbf{K}$ belong to $\mathcal{C}$.

Consider a language L_1 and a *finite multiset of languages* $\mathbf{L}_2$ (we speak of "multiset" here for the sake of allowing several copies of the same language in $\mathbf{L}_2$ which will be convenient for technical reasons). A *separating cover* for the pair $(L_1, \mathbf{L}_2)$ is a cover $\mathbf{K}$ of L_1 such that for any $K \in \mathbf{K}$, there exists $L \in \mathbf{L}_2$ which satisfies $K \cap L = \emptyset$. Finally, given a class $\mathcal{C}$, we say that the pair $(L_1, \mathbf{L}_2)$ is $\mathcal{C}$-coverable when there exists a separating $\mathcal{C}$-cover. The $\mathcal{C}$-covering problem is now defined as follows:

INPUT: A regular language L_1 and a finite multiset of regular languages $\mathbf{L}_2$.
OUTPUT: Is $(L_1, \mathbf{L}_2)$ $\mathcal{C}$-coverable ?

Let us complete this definition by explaining why covering generalizes separation: the latter is special case of the former when the multiset $\mathbf{L}_2$ is a singleton.

Fact 2.2. *Let $\mathcal{C}$ be a lattice and L_1, L_2 two languages. Then L_1 is $\mathcal{C}$-separable from L_2, if and only if $(L_1, \{L_2\})$ is $\mathcal{C}$-coverable.*

Proof. Assume first that L_1 is $\mathcal{C}$-separable from L_2 and let $K \in \mathcal{C}$ be a separator. Then, it is immediate that $\mathbf{K} = \{K\}$ is a separating $\mathcal{C}$-cover for $(L_1, \{L_2\})$. Conversely, assume that $(L_1, \{L_2\})$ is $\mathcal{C}$-coverable and let $\mathbf{K}$ be a separating $\mathcal{C}$-cover. Moreover, let K be the union of all languages in $\mathbf{K}$. We have $K \in \mathcal{C}$ since $\mathcal{C}$ is a lattice. Clearly, $L_1 \subseteq K$ since $\mathbf{K}$ was a cover of L_1. Moreover, we know that no language in $\mathbf{K}$ intersects L_2 since $\mathbf{K}$ was separating. Thus, $L_2 \cap K = \emptyset$ which means that $K \in \mathcal{C}$ separates L_1 from L_2. $\square$

Remark 2.3. *While covering is a natural generalization of separation, we actually consider it out of necessity. For the classes that we investigate in the paper, no "direct" separation algorithm is known. Our techniques require considering the covering problem (we come back to this point in Section 5).*

2.3. **Finite lattices and stratifications.**

All classes that we consider in the paper are built from an arbitrary finite lattice (i.e. one that contains finitely many languages) using a generic construction process (presented in Section 3). Consequently, finite lattices will be important in the paper. It turns out that they have several convenient properties that we present here.

Canonical preorder relations. Consider a finite lattice $\mathcal{C}$. One may associate a *canonical preorder relation over* A^* to $\mathcal{C}$. The definition is as follows. Given $w, w' \in A^*$, we write $w \leqslant_{\mathcal{C}} w'$ if and only if the following holds:

$$\text{For all } L \in \mathcal{C}, \quad w \in L \Rightarrow w' \in L.$$

It is immediate from the definition that $\leqslant_{\mathcal{C}}$ is transitive and reflexive, making it a preorder.

Example 2.4. *We let* AT *as the class of languages consisting of all Boolean combinations of languages* A^*aA^**, for some* $a \in A$*. Though this terminology is not standard, "*AT*" stands for "alphabet testable":* $L \in \mathrm{AT}$ *if and only if membership of a word* w *in* L *depends only on the set of letters occurring in* w*. Clearly,* AT *is a finite Boolean algebra. In that case,* $\leqslant_{\mathrm{AT}}$ *is an equivalence relation which we denote by* $\sim_{\mathrm{AT}}$*: one may verify that* $w \sim_{\mathrm{AT}} w'$ *if and only if* w *and* w' *have the same alphabet (i.e. contain the same set of letters).*

The relation $\leqslant_{\mathcal{C}}$ has many applications. We start with an important lemma, which relies on the fact that $\mathcal{C}$ is finite. We say that a language $L \subseteq A^*$ is an *upper set* (for $\leqslant_{\mathcal{C}}$) when for any two words $u, v \in A^*$, if $u \in L$ and $u \leqslant_{\mathcal{C}} v$, then $v \in L$.

Lemma 2.5. *Let* $\mathcal{C}$ *be a finite lattice. Then, for any* $L \subseteq A^*$*, we have* $L \in \mathcal{C}$ *if and only if* L *is an upper set for* $\leqslant_{\mathcal{C}}$*. In particular,* $\leqslant_{\mathcal{C}}$ *has finitely many upper sets.*

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

$$L = \bigcup_{w \in L} \uparrow w.$$

Hence, since $\mathcal{C}$ is closed under union and is finite, L belongs to $\mathcal{C}$. ☐

While Lemma 2.5 states an equivalence, we mainly use the left to right implication (or rather its contrapositive). One may apply it to show that a language does not belong to $\mathcal{C}$. Indeed, by the lemma, proving that $L \notin \mathcal{C}$ is the same as proving that L is not an upper set for $\leqslant_{\mathcal{C}}$. In other words, one needs to exhibit $u, v \in A^*$ such that $u \leqslant_{\mathcal{C}} v$, $u \in L$ and $v \notin L$.

Example 2.6. *Assume that* $A = \{a, b\}$ *and consider the class* AT *of Example 2.4. The language* $L = A^*aA^*bA^*$ *does not belong to* AT*. Indeed,* $ab \sim_{\mathrm{AT}} ba$*,* $ab \in L$ *and* $ba \notin L$*.*

Let us complete these definitions with a few additional useful results. First, as we observed for AT in Example 2.4, when the finite lattice $\mathcal{C}$ is actually a Boolean algebra, it turns out that $\leqslant_{\mathcal{C}}$ is an equivalence relation, which we shall denote by $\sim_{\mathcal{C}}$.

Lemma 2.7. *Let* $\mathcal{C}$ *be a finite Boolean algebra. Then for any alphabet* A*, the canonical preorder* $\leqslant_{\mathcal{C}}$ *is an equivalence relation* $\sim_{\mathcal{C}}$ *which admits the following direct definition :*

$$w \sim_{\mathcal{C}} w' \text{ if and only if for all } L \in \mathcal{C}, \quad w \in L \Leftrightarrow w' \in L$$

Thus, for any $L \subseteq A^*$*, we have* $L \in \mathcal{C}$ *if and only if* L *is a union of* $\sim_{\mathcal{C}}$*-classes. In particular,* $\sim_{\mathcal{C}}$ *has finite index.*

Proof. It is clear that when $w \sim_{\mathcal{C}} w'$, we have $w \leqslant_{\mathcal{C}} w'$ as well. We prove the reverse implication. Let $w, w' \in A^*$ be such that $w \leqslant_{\mathcal{C}} w'$. Let $L \in \mathcal{C}$ and observe that by closure under complement, we know that $A^* \setminus L \in \mathcal{C}$. Therefore, by definition of $w \leqslant_{\mathcal{C}} w'$,

$$\begin{array}{lll} w \in L & \Rightarrow & w' \in L \\ w \in A^* \setminus L & \Rightarrow & w' \in A^* \setminus L \end{array}$$

One may now combine the first implication with the contrapositive of the second, which yields $w \in L \Leftrightarrow w' \in L$. We conclude that $w \sim_{\mathcal{C}} w'$: $\leqslant_{\mathcal{C}}$ and $\sim_{\mathcal{C}}$ are the same relation. □

Another important and useful property is that when $\mathcal{C}$ is quotienting, the canonical preorder $\leqslant_{\mathcal{C}}$ is compatible with word concatenation.

Lemma 2.8. *A finite lattice $\mathcal{C}$ is quotienting if and only if its associated canonical preorder $\leqslant_{\mathcal{C}}$ is compatible with word concatenation. That is, for any words u, v, u', v',*

$$u \leqslant_{\mathcal{C}} u' \quad \text{and} \quad v \leqslant_{\mathcal{C}} v' \quad \Rightarrow \quad uv \leqslant_{\mathcal{C}} u'v'.$$

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

Conversely, assume that $\leqslant_{\mathcal{C}}$ is a precongruence. Let $L \in \mathcal{C}$ and $w \in A^*$, we prove that $w^{-1}L \in \mathcal{C}$ (the proof for right quotients is symmetrical). By Lemma 2.5, we have to prove that $w^{-1}L$ is an upper set. Let $u \in w^{-1}L$ and $u' \in A^*$ such that $u \leqslant_{\mathcal{C}} u'$. Since $\leqslant_{\mathcal{C}}$ is a precongruence, we have $wu \leqslant_{\mathcal{C}} wu'$. Hence, since L is an upper set (it belongs to $\mathcal{C}$) and $wu \in L$, we have $wu' \in L$. We conclude that $u' \in w^{-1}L$, which terminates the proof. □

Stratifications. While the above notions are useful, the downside is that they only apply to *finite* lattices. However, it is possible to lift their benefits to infinite classes with the notion of *stratification*. Consider an arbitrary *infinite* quotienting lattice $\mathcal{C}$. A *stratification* of $\mathcal{C}$ is an infinite sequence $\mathcal{C}_0, \dots, \mathcal{C}_k, \dots$ of *finite* quotienting lattices such that,

$$\text{For all } k, \quad \mathcal{C}_k \subseteq \mathcal{C}_{k+1} \quad \text{and} \quad \mathcal{C} = \bigcup_{k \in \mathbb{N}} \mathcal{C}_k.$$

The point here is once we have a stratification of $\mathcal{C}$ in hand, we may now associate a canonical preorder $\leqslant_k$ to each stratum $\mathcal{C}_k$. Proving that a language L does not belong to $\mathcal{C}$ now amounts to proving that it does not belong any stratum $\mathcal{C}_k$: for all $k \in \mathbb{N}$, one needs to exhibit $u, v \in A^*$ such that $u \leqslant_k v$, $u \in L$ and $v \notin L$.

3. Concatenation hierarchies and main theorem

In this section, we present the classes that we intend to investigate. We actually consider a family of classes which are all *built using a simple generic construction process.* As we shall see, this family includes several prominent classes coming from mathematical logic.

We start by defining the family of classes that we consider. Then, we present the main theorem of the paper: covering (and therefore separation as well) is decidable for any class in the family. We illustrate this theorem by discussing its consequences and presenting the most prominent classes to which it applies.

3.1. **Closure operations.** Our generic theorem states that covering is decidable for any class which is built from a **finite** quotienting Boolean algebra using a generic construction process that we present now. This construction is based on two operations that one may apply to classes of languages: *Boolean closure and polynomial closure.*

Boolean closure. As the name, suggests, the Boolean closure of a class $\mathcal{C}$ (denoted by $Bool(\mathcal{C})$) is simply the smallest Boolean algebra which contains $\mathcal{C}$. Observe that we have the following lemma.

Lemma 3.1. *For any quotienting lattice $\mathcal{C}$, its Boolean closure $Bool(\mathcal{C})$ is a quotienting Boolean algebra.*

Proof. This is immediate: quotients commute with Boolean operations. □

Polynomial closure. Consider a class $\mathcal{C}$, its *polynomial closure* (denoted by $Pol(\mathcal{C})$) is the smallest *lattice* containing $\mathcal{C}$ and *closed under marked concatenation*:

$$\text{For all } K, L \in Pol(\mathcal{C}) \text{ and all } a \in A, \quad KaL \in Pol(\mathcal{C}).$$

We have the following simple lemma which will be useful.

Lemma 3.2. *For any quotienting lattice $\mathcal{C}$, its polynomial closure $Pol(\mathcal{C})$ is a quotienting lattice closed under concatenation and marked concatenation.*

Proof. It is immediate by definition that $Pol(\mathcal{C})$ is a lattice closed under marked concatenation. Hence, we focus on closure under quotients and classical concatenation. Let us start with quotients. Let $L \in Pol(\mathcal{C})$ and $w \in A^*$. We have to prove that $w^{-1}L \in Pol(\mathcal{C})$ and $Lw^{-1} \in Pol(\mathcal{C})$. By symmetry, we concentrate on the former. Moreover, it is simple to verify that given $b \in A$ and $u \in A^*$, we have $(ub)^{-1}L = b^{-1}(u^{-1}L)$. Thus, it follows from a simple induction that it suffices to consider the case when w is a single letter $b \in A$.

By definition of $Pol(\mathcal{C})$, the language L is built from languages of $\mathcal{C}$ using a finite number of times union, intersection and marked concatenation. We argue by induction on this construction. If L belongs to $\mathcal{C}$, then so does $b^{-1}L$ since $\mathcal{C}$ is closed under quotients, so in particular $b^{-1}L \in Pol(\mathcal{C})$. If $L = K \cup H$ with $K, H \in Pol(\mathcal{C})$, then $b^{-1}L = (b^{-1}K) \cup (b^{-1}H)$. By induction, both $b^{-1}K$ and $b^{-1}H$ belong to $Pol(\mathcal{C})$, and since $Pol(\mathcal{C})$ is closed under union, $b^{-1}L$ also belongs to $Pol(\mathcal{C})$. If $L = K \cap H$ with $K, H \in Pol(\mathcal{C})$, we argue similarly using $b^{-1}L = (b^{-1}K) \cap (b^{-1}H)$. Finally, if $L = KaH$, then we have,

$$b^{-1}L = \begin{cases} (b^{-1}K)aH \cup H & \text{if } \varepsilon \in H \text{ and } a = b \\ (b^{-1}K)aH & \text{otherwise} \end{cases}$$

By induction, the language $b^{-1}K$ belong to $Pol(\mathcal{C})$, hence so does L since $Pol(\mathcal{C})$ is closed under union and marked concatenation.

It remains to show that $Pol(\mathcal{C})$ is closed under concatenation. Let $K, L \in Pol(\mathcal{C})$. One may verify that,

$$KL = \begin{cases} K \cup \bigcup_{a \in A} Ka(a^{-1}L) & \text{if } \varepsilon \in L, \\ \bigcup_{a \in A} Ka(a^{-1}L) & \text{if } \varepsilon \notin L. \end{cases}$$

Since we already know that $Pol(\mathcal{C})$ is closed under union, quotients and marked concatenation, we obtain $KL \in Pol(\mathcal{C})$, which concludes the proof. □

For the sake of avoiding clutter, we shall write $BPol(\mathcal{C})$ for $Bool(Pol(\mathcal{C}))$ and $PBPol(\mathcal{C})$ for $Pol(BPol(\mathcal{C}))$. It was shown in [PZ17c] that for any **finite** quotienting Boolean algebra $\mathcal{C}$, separation[1] is decidable for both $Pol(\mathcal{C})$ and $BPol(\mathcal{C})$.

Our main theorem in the paper extends these results: we show that for any **finite** quotienting Boolean algebra $\mathcal{C}$, separation and covering are decidable for the class $PBPol(\mathcal{C})$. Moreover, we also present a new proof for the decidability of $Pol(\mathcal{C})$-covering. Note that we do this out of necessity: we actually prove a result for $Pol(\mathcal{C})$ which is stronger than the decidability of covering and that we require for handling $PBPol(\mathcal{C})$).

Remark 3.3. *The arguments used in the paper for handling $PBPol(\mathcal{C})$ are mostly independent from the ones used in* [PZ17c] *for $BPol(\mathcal{C})$. The key idea is that both results build on our knowledge of the simpler class $Pol(\mathcal{C})$. However, they do so in orthogonal directions.*

Before we state our main theorem properly, let us finish the definitions with two useful results about classes of the form $PBPol(\mathcal{C})$ that we shall need later. A first result is that such classes always contain all finite languages.

Lemma 3.4. *Let $\mathcal{C}$ be an arbitrary lattice. Then, $BPol(\mathcal{C})$ (and therefore $PBPol(\mathcal{C})$ as well) contains all finite languages.*

Proof. Since $BPol(\mathcal{C})$ is closed under union, it suffices to show that for any $w \in A^*$, the singleton $\{w\}$ belongs to $BPol(\mathcal{C})$. Let $w = a_1 \cdots a_n$ be the decomposition of w as a concatenation of letters. Since $\mathcal{C}$ is a lattice, we have $A^* \in \mathcal{C}$. Thus, $L = A^*a_1A^*a_2 \cdots A^*a_nA^* \in Pol(\mathcal{C})$. Moreover, for any $b \in A$, we have $LbA^* \in Pol(\mathcal{C})$. Therefore, we have,

$$\{w\} = L \setminus \left(\bigcup_{b \in A} LbA^* \right) \in BPol(\mathcal{C})$$

This concludes the proof. $\square$

The second result is used to bypass Boolean closure in the definition of $PBPol(\mathcal{C})$. Given any class $\mathcal{D}$, we write $\overline{\mathcal{D}}$ the *complement class* of $\mathcal{D}$. That is, $\overline{\mathcal{D}}$ contains all languages of the form $A^* \setminus L$ where $L \in \mathcal{D}$. One may verify the following fact.

Fact 3.5. *For any quotienting lattice $\mathcal{D}$, the complement class $\overline{\mathcal{D}}$ is a quotienting lattice.*

We may now state our second result: Boolean closure may be replaced by complement in the definition of $PBPol(\mathcal{C})$.

Lemma 3.6. *Let $\mathcal{C}$ be some class. Then, $PBPol(\mathcal{C}) = Pol(\overline{Pol(\mathcal{C})})$.*

Proof. By definition, it is clear that $\overline{Pol(\mathcal{C})} \subseteq BPol(\mathcal{C})$. Hence, the right to left inclusion $Pol(\overline{Pol(\mathcal{C})}) \subseteq PBPol(\mathcal{C})$ is immediate and we may concentrate on the converse one.

We show that $BPol(\mathcal{C}) \subseteq Pol(\overline{Pol(\mathcal{C})})$. It will then follow that we have $PBPol(\mathcal{C}) \subseteq Pol(Pol(\overline{Pol(\mathcal{C})})) = Pol(\overline{Pol(\mathcal{C})})$ as desired. Let L be a language in $BPol(\mathcal{C})$. By definition, L is a Boolean combination of languages in $Pol(\mathcal{C})$. Therefore, it follows from DeMorgan's laws that L is built by applying unions and intersections to languages that are either in $Pol(\mathcal{C})$ or in $\overline{Pol(\mathcal{C})}$. It follows that $L \in Pol(\overline{Pol(\mathcal{C})})$, since $Pol(\mathcal{C}) \subseteq Pol(\overline{Pol(\mathcal{C})})$, $\overline{Pol(\mathcal{C})} \subseteq Pol(\overline{Pol(\mathcal{C})})$ and $Pol(\overline{Pol(\mathcal{C})})$ is closed under union and intersection. $\square$

[1] In fact, while this is not explicitly stated in the paper the arguments of [PZ17c] apply to covering.

3.2. **Main theorem.** We are now ready to state the main theorem of the paper: $PBPol(\mathcal{C})$-covering is decidable when $\mathcal{C}$ is a finite quotienting Boolean algebra.

Theorem 3.7. *Let $\mathcal{C}$ be a finite quotienting Boolean algebra. Then covering and separation are both decidable for $Pol(\mathcal{C})$ and $PBPol(\mathcal{C})$.*

All remaining sections are devoted to the proof of Theorem 3.7. We use Section 4 to introduce some mathematical tools that we shall need in our proofs: Simon's factorization theorem [Sim90, Kuf08] and a generic stratification for $Pol(\mathcal{C})$ when $\mathcal{C}$ is a finite quotienting lattice. Then, we devote Sections 5 and 6 to presenting a general framework which is designed for handling covering problems (it was originally introduced in [PZ16, PZ17b]). We shall use this framework to formulate our two covering algorithms. Finally, Section 7 and 8 are devoted to presenting and proving our covering algorithms for $Pol(\mathcal{C})$ and $PBPol(\mathcal{C})$ respectively. However, before we turn to the proof of Theorem 3.7, let us discuss its applications.

3.3. **Concatenation hierarchies.** The polynomial and Boolean closure operations are important as they are involved in the definition of natural hierarchies of classes of languages: *concatenation hierarchies*. Here, we briefly recall what they are (we refer the reader to [Pin98, PW97, Pin17, Str88] for details). Such a hierarchy depends on a single parameter: a quotienting Boolean algebra of regular languages $\mathcal{C}$, called its *basis*. Once the basis is chosen, the construction is uniform. Languages are classified into levels of two kinds: full levels (denoted by $0, 1, 2, \dots$) and half levels (denoted by $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots$):

- Level 0 is the basis (*i.e.*, our parameter class $\mathcal{C}$).
- Each *half level* $n + \frac{1}{2}$, for $n \in \mathbb{N}$, is the *polynomial closure* of the previous full level, *i.e.*, of level n.
- Each *full level* $n+1$, for $n \in \mathbb{N}$, is the *Boolean closure* of the previous half level, *i.e.*, of level $n + \frac{1}{2}$.

$$0 \xrightarrow{Pol} \tfrac{1}{2} \xrightarrow[Bool]{} 1 \xrightarrow{Pol} \tfrac{3}{2} \xrightarrow[Bool]{} 2 \xrightarrow{Pol} \tfrac{5}{2} \cdots\cdots$$

Figure 1: A concatenation hierarchy

Therefore, Theorem 3.7 may be reformulated as follows: if $\mathcal{C}$ is a **finite** basis, separation and covering are decidable for the levels $\frac{1}{2}$ and $\frac{3}{2}$ in the associated concatenation hierarchy. There are two prominent examples of concatenation hierarchies with a finite basis:

- The *Straubing-Thérien* hierarchy [Str81, Thé81], its basis is $\{\emptyset, A^*\}$.
- The *dot-depth* hierarchy of Brzozowski and Cohen [BC71], its basis is $\{\emptyset, \{\varepsilon\}, A^+, A^*\}$.

These two hierarchies are strict [BK78]. Solving separation (and membership) for all their levels is a longstanding open problem. It was already known that separation is decidable for the levels $\frac{1}{2}$, 1, $\frac{3}{2}$ and 2 for both hierarchies [CMM13, PvRZ13, PZ14, PZ17c, PZ15a]. However, it turns out that for these two particular hierarchies, we are able to lift our results to the level $\frac{7}{2}$.

Recall that AT denotes the class of alphabet testable languages: it consists of all Boolean combinations of languages A^*aA^*, for some letter $a \in A$ (see Example 2.4). The following lemma was shown in [PS85].

Lemma 3.8. *The level* 2 *in the Straubing-Thérien hierarchy is exactly the class* $BPol(\text{AT})$.

Since the level $\frac{7}{2}$ of the Straubing-Thérien hierarchy is the polynomial closure of the level 2 by definition, it is immediate from Lemma 3.8 that it is exactly the class $PBPol(\text{AT})$. Therefore, since it is straightforward to verify that AT is a finite quotienting Boolean algebra, the following corollary is an immediate consequence of Theorem 3.7.

Corollary 3.9. *The separation and covering problems are decidable for the level* $\frac{7}{2}$ *of the Straubing-Thérien hierarchy*

Additionally, it was shown in [PZ14], that for any **half** level n in Straubing-Thérien hierarchy, membership for the level $n+1$ reduces to separation for the level n. Therefore, we also obtain the following immediate corollary.

Corollary 3.10. *It decidable to test whether a regular language belongs to the level* $\frac{9}{2}$ *in the Straubing-Thérien hierarchy*

Finally, it is known that these two results may be lifted to the dot-depth hierarchy using generic transfer results (see [Str85] for membership and [PZ15a] for separation). Therefore, we also get the two following additional results.

Corollary 3.11. *The separation problem is decidable for the level* $\frac{7}{2}$ *of the dot-depth hierarchy*

Corollary 3.12. *It decidable to test whether a regular language belongs to the level* $\frac{9}{2}$ *in the dot-depth hierarchy*

Remark 3.13. *There is no published transfer result for covering which explains why we do not discuss it for the dot-depth hierarchy. However, while this is not explicitly stated in* [PZ15a]*, the techniques used for handling obtaining a separation transfer result apply to covering as well. Therefore, the decidability of covering for dot-depth* $\frac{7}{2}$ *may also be obtained from our results.*

3.4. **First-order logic and quantifier alternation.** We finish the section by discussing the connection with quantifier alternation hierarchies.

Let us briefly recall the definition of first-order logic over words. One may view a finite word as a logical structure made of a sequence of positions. Each position carries a label in the alphabet A and can be quantified. We denote by "$<$" the linear order over the positions. We consider first-order logic, which is equipped with the following signature:

- For each $a \in A$, a unary predicate P_a selecting positions labeled with the letter "a".
- A binary predicate "$<$" interpreted as the linear order.
- A binary predicate "$+1$" interpreted as the successor relation.
- Unary predicates "min" and "max" selecting the leftmost and rightmost positions.
- A constant "ε" which holds when the word is empty.

We shall write $\text{FO}(<, +1, min, max, \varepsilon)$ for this variant of first-order (we consider another variant with a restricted signature below). For every sentence φ, one may associate the language $\{w \in A^* \mid w \models \varphi\}$ of words satisfying φ. Therefore, $\text{FO}(<, +1, min, max, \varepsilon)$ defines the class of languages that can be defined with an $\text{FO}(<, +1, min, max, \varepsilon)$ sentence. It is known that this class is a quotienting Boolean algebra.

Here, we are not interested in first-order logic itself: we consider its quantifier alternation hierarchy. One may classify sentences by counting their number of quantifier alternations. Let $n \in \mathbb{N}$. We say that a sentence is $\Sigma_n(<, +1, min, max, \varepsilon)$ (resp. $\Pi_n(<, +1, min, max, \varepsilon)$) when its prenex normal form has either,

- *exactly* $n-1$ quantifier alternations (*i.e.*, exactly n blocks of quantifiers) starting with an $\exists$ (resp. $\forall$), or
- *strictly less* than $n-1$ quantifier alternations (*i.e.*, strictly less than n blocks of quantifiers).

For example, a formula whose prenex normal form is

$$\forall x_1 \exists x_2 \forall x_3 \forall x_4 \ \varphi(x_1, x_2, x_3, x_4) \quad \text{(with } \varphi \text{ quantifier-free)}$$

is $\Pi_3(<, +1, min, max, \varepsilon)$. In general, the negation of a $\Sigma_n(<, +1, min, max, \varepsilon)$ sentence is not a $\Sigma_n(<, +1, min, max, \varepsilon)$ sentence (it is a $\Pi_n(<, +1, min, max, \varepsilon)$ sentence). Hence it is relevant to define $\mathcal{B}\Sigma_n(<, +1, min, max, \varepsilon)$ sentences as the Boolean combinations of $\Sigma_n(<, +1, min, max, \varepsilon)$ sentences. This gives a hierarchy of classes of languages as presented in Figure 2.

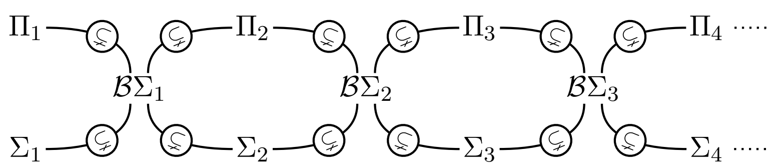

Figure 2: Quantifier Alternation Hierarchy

It was shown by Thomas [Tho82] that this hierarchy corresponds exactly to the dot-depth hierarchy of Brzozowski and Cohen. For any $n \in \mathbb{N}$, dot-depth n corresponds to $\mathcal{B}\Sigma_n(<, +1, min, max, \varepsilon)$ and dot-depth $n + \frac{1}{2}$ to $\Sigma_{n+1}(<, +1, min, max, \varepsilon)$. Therefore, Corollaries 3.11 and 3.12 may be translated as follows from a logical point of view.

Corollary 3.14. *The separation problem is decidable for* $\Sigma_3(<, +1, min, max, \varepsilon)$.

Corollary 3.15. *It decidable to test whether a regular language may be defined by a sentence of* $\Sigma_4(<, +1, min, max, \varepsilon)$.

Finally, it is possible to adapt the result of Thomas to obtain a similar correspondence between the Straubing-Thérien hierarchy and the quantifier alternation hierarchy within a variant of first-order logic using a smaller signature. We denote by FO($<$) the variant of first-order logic whose signature only contains the label predicates and the linear order (i.e. $+1, min, max$ and ε are disallowed).

Remark 3.16. *This restriction does do change the expressive power of first-order logic as a whole:* $+1, min, max$ *and* ε *can be defined from "*$<$*". For example, the formula* $y = x + 1$ *is equivalent to* $(x < y) \wedge \neg(\exists z \ x < z < y)$. *However, this definition costs quantifier alternations. Thus,* FO($<$) *and* FO($<, +1, min, max, \varepsilon$) *have different quantifier alternation hierarchies.*

One may define a quantifier alternation hierarchy for FO($<$) in the natural way. We shall denote by $\Sigma_n(<)$, $\Pi_n(<)$ and $\mathcal{B}\Sigma_n(<)$ its levels. It was shown by Perrin and Pin [PP86] that this second alternation hierarchy corresponds exactly to the Straubing-Thérien hierarchy. Therefore, Corollaries 3.9 and 3.10 may be translated as follows from a logical point of view.

Corollary 3.17. *The separation and covering problems are decidable for* $\Sigma_3(<)$*.*

Corollary 3.18. *It decidable to test whether a regular language is definable in* $\Sigma_4(<)$*.*

Remark 3.19. *The two above correspondences between alternation hierarchies and concatenation hierarchies are not coincidental. It turns out that given any basis* $\mathcal{C}$*, one may define a set of first-order predicates* S *such that the concatenation hierarchy of basis* $\mathcal{C}$ *corresponds exactly to the quantifier alternation hierarchy within the variant of first-order logic whose signature is* S *(see* [PZ17a, PZ17c]*).*

4. Mathematical tools

In this section, we present two mathematical tools which we shall use later when proving the correction of our covering algorithms. These two objects are independent. First, we present the *factorization forest theorem of Simon*. Then, we present a generic stratification for classes of the form $Pol(\mathcal{C})$ when $\mathcal{C}$ is a finite quotienting lattice.

4.1. Factorization forests.

Simon's factorization theorem is a combinatorial result which applies to *finite semigroups*. Here, we briefly recall this theorem. We refer the reader to [Kuf08, Boj09, Col10] for more details and a proof.

A *semigroup* is a set S equipped with an associative multiplication (usually denoted by "$\cdot$"). Observe that A^+ is a semigroup when equipped with word concatenation as the multiplication. Recall that an idempotent within a semigroup S is an element $e \in S$ such that $ee = e$. It is well-known that when a semigroup S is *finite*, there exists a number $\omega(S)$ (denoted by ω when S is understood from the context) such that for each element $s \in S$, s^ω is an idempotent. Finally, a *monoid* M is a semigroup in which there exists a neutral element denoted 1_M. Observe that A^* is a monoid with concatenation as the multiplication and ε as the neutral element.

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

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

Given a word $w \in A^*$, an *α-factorization forest for w* is an α-factorization forest whose root is labeled by w. Moreover, the *height* of a factorization forest is the largest $h \in \mathbb{N}$ such that it contains a branch with h inner nodes (i.e. a single leaf has height 0). Additionally, we shall consider the notion of *idempotent height*. The idempotent height of an α-factorization forest is the largest number $p \in \mathbb{N}$ such that there exists a branch containing p idempotent nodes in the forest. We often use the following notations. Let $s \in M$ and any $h, p \in \mathbb{N}$.

(1) $F^\alpha(s,h,p) \subseteq A^*$ denotes the language of all words $w \in \alpha^{-1}(s)$ admitting an α-factorization forest of height at most h and idempotent height at most p.
(2) $F^\alpha_B(s,h,p) \subseteq A^*$ denotes the language of all words $w \in \alpha^{-1}(s)$ admitting an α-factorization forest of height of at most h, of idempotent height at most p and whose root is a *binary node*.
(3) $F^\alpha_I(s,h,p) \subseteq A^*$ denotes the language of all words $w \in \alpha^{-1}(s)$ admitting an α-factorization forest of height of at most h, of idempotent height at most p and whose root is a *idempotent node*.

We turn to Simon's factorization forest theorem: there exists a bound depending only on M such that any word admits an α-factorization forest of height smaller than this bound.

Theorem 4.1 ([Sim90, Kuf08])**.** *Consider a morphism $\alpha : A^* \to M$. For all words $w \in A^*$, there exists an α-factorization forest for w of height smaller than $3|M|-1$.*

Observe that an immediate consequence of Theorem 4.1 is that for any $s \in M$, we have $\alpha^{-1}(s) = F^\alpha(s, 3|M|-1, 3|M|-1)$. We shall use Theorem 4.1 conjointly with a second result on factorization forests which takes the idempotent height into account. Given a word $w \in A^*$, an *infix* of w is a second word $u \in A^*$ such that $w = v_1uv_2$ for some $v_1, v_2 \in A^*$.

Proposition 4.2. *Consider a morphism $\alpha : A^* \to M$ and let $p, h \in \mathbb{N}$. Let $w \in A^*$ admitting an α-factorization forest of height h and idempotent height p. Then any infix of w admits an α-factorization forest of height at most $h+2$ and idempotent height at most p.*

Let us prove Proposition 4.2. The argument is an induction on the α-factorization forest of w: we show that it may be "repaired" into another forest for some infix of w without adding any idempotent node. We first consider the special case of prefixes and suffixes.

Lemma 4.3. *Let $p, h \in \mathbb{N}$ and consider a word $w \in A^*$ admitting an α-factorization forest of height h and idempotent height p. Then any prefix or suffix of w admits an α-factorization forest of height at most $h+1$ and idempotent height at most p.*

Observe that any infix of a word $w \in A^*$ is by definition the prefix of a suffix of w. Hence, Proposition 4.2 is obtained by applying Lemma 4.3 twice (once for prefixes and once for suffixes). Therefore, we may concentrate on proving Lemma 4.3. We only treat the case of prefixes (the argument for suffixes is symmetrical).

Let $w \in A^*$ admitting an α-factorization forest of height h and idempotent height p. We shall denote this forest by $\mathcal{F}$. Consider a prefix u of w. We construct an α-factorization forest for u using structural induction on $\mathcal{F}$. There are three cases depending on the root node of $\mathcal{F}$. If the root of $\mathcal{F}$ is a leaf node, then $w = a$ or $w = \varepsilon$. In particular, $p = h = 0$. Since u is a prefix of w, we have $u = a$ or $u = \varepsilon$. Hence u admits an α-factorization forest of height $0 = h$ and idempotent height $0 = p$.

If the root of $\mathcal{F}$ is a binary node, it has two children labeled with words $w_1, w_2 \in A^*$ such that $w = w_1w_2$. Moreover, w_1, w_2 admit α-factorization forests of heights at most $h-1$ and idempotent heights at most p. Since u is a prefix of $w = w_1w_2$ by hypothesis, there are two possibles cases. First, u may be a prefix of w_1. In this case, it suffices to apply induction to the forest of w_1 to get the desired forest for u. Otherwise, there exists a prefix u' of w_2 such that $u = w_1u'$. We may apply induction to the forest of w_2 to get a forest of height at most $h-1+1 = h$ and idempotent height at most p for u'. Using one binary node, one may then combine the forests of w_1 and u' into a single forest for $u = w_1u'$. By construction, this new forest has height at most $h+1$ and idempotent height at most p.

We finish with the case when the root of $\mathcal{F}$ is an idempotent node. Its children are labeled with words $w_1, \dots, w_n$ such that $w = w_1 \cdots w_n$ and $\alpha(w_1) = \cdots = \alpha(w_n) = e$ for some idempotent $e \in M$. Moreover, the words $w_1, \cdots, w_n$ admit α-factorization forests of heights at most $h-1$ and idempotent heights at most $p-1$. Let i be the smallest natural such that u is a prefix of $w_1 \cdots w_i$. If $i = 1$, u is a prefix of w_1 and it suffices to apply induction to the forest of w_1 to get the desired forest for u. Otherwise, there exists a prefix u' of w_i such that $u = w_1 \cdots w_{i-1} u'$. We know that,

- u' admits an α-factorization forest of height at most $(h-1)+1 = h$ and idempotent height at most $p-1$ (this is by induction).
- $w_1 \cdots w_{i-1}$ admit an α-factorization forest of height at most h and idempotent height at most p (the root is an idempotent node whose children are labeled with $w_1, \dots, w_{i-1}$).

These two forests can be combined into a single forest for u with one binary node. By definition, this forest has height at most $h+1$ and idempotent height at most p.

4.2. **A stratification for** $Pol(\mathcal{C})$**.** We turn to the second mathematical tool that we shall need: a stratification for classes of the form $Pol(\mathcal{C})$ when $\mathcal{C}$ is a finite quotienting lattice.

Remark 4.4. *We do not present a stratification for $PBPol(\mathcal{C})$ (even if our ultimate objective is to solve $PBPol(\mathcal{C})$-covering). Indeed, it turns out that the stratification of $Pol(\mathcal{C})$ that we present here suffices to handle both $Pol(\mathcal{C})$ and $PBPol(\mathcal{C})$.*

We fix an arbitrary finite quotienting lattice $\mathcal{C}$ and define a finite quotienting lattice $Pol_k(\mathcal{C})$ for each $k \in \mathbb{N}$. The definition uses induction on k and counts the number of marked concatenations that are necessary to define each language in $Pol(\mathcal{C})$.

- When $k = 0$, we simply define $Pol_0(\mathcal{C}) = \mathcal{C}$.
- When $k \geq 1$, $Pol_k(\mathcal{C})$ is the smallest lattice such that:
  1. $Pol_{k-1}(\mathcal{C}) \subseteq Pol_k(\mathcal{C})$.
  2. For any $a \in A$ and $L_1, L_2 \in Pol_{k-1}(\mathcal{C})$, we have $L_1 a L_2 \in Pol_k(\mathcal{C})$.

This concludes the definition. Since $\mathcal{C}$ was a finite lattice, it is immediate that all classes $Pol_k(\mathcal{C})$ are finite lattices as well. Moreover, we indeed have,

$$\text{For all } k \in \mathbb{N},\ Pol_k(\mathcal{C}) \subseteq Pol_{k+1}(\mathcal{C}) \quad \text{and} \quad Pol(\mathcal{C}) = \bigcup_{k \in \mathbb{N}} Pol_k(\mathcal{C})$$

Finally, it is straightforward to verify that all classes $Pol_k(\mathcal{C})$ are quotienting (the argument is identical to the one used in Lemma 3.2 for proving that the whole class $Pol(\mathcal{C})$ is quotienting). Thus, we did define a stratification of $Pol(\mathcal{C})$. We now prove a few properties of this stratification that we shall need.

Consider the canonical preorder relations associated to the strata. For any $k \in \mathbb{N}$, we denote by $\leqslant_k$ the preorder associated to $Pol_k(\mathcal{C})$. Recall that for any $w, w' \in A^*$, we have,

$$w \leqslant_k w' \quad \text{if and only if} \quad \text{for all languages } L \in Pol_k(\mathcal{C}),\ w \in L \Rightarrow w' \in L$$

We showed in Lemma 2.5 that the languages in $Pol_k(\mathcal{C})$ are exactly the upper sets for the relation $\leqslant_k$. Moreover, since the lattices $Pol_k(\mathcal{C})$ are quotienting, we know from Lemma 2.8 that the relations $\leqslant_k$ are compatible with word concatenation.

We present two specific properties of the preorders $\leqslant_k$. We start with an alternate definition of $\leqslant_k$ which is easier to manipulate when proving these properties. Recall that we write $\leqslant_\mathcal{C}$ the canonical preorder associated to the finite quotienting lattice $\mathcal{C}$.

Lemma 4.5. *Let A be an alphabet and $k \in \mathbb{N}$. For any $w, w' \in A^*$, we have $w \leqslant_k w'$ if and only if the two following properties hold:*

(1) $w \leqslant_\mathcal{C} w'$
(2) If $k > 0$, for any decomposition $w = uav$ with $u, v \in A^*$ and $a \in A$, there exist $u', v' \in A^*$ such that $w' = u'av'$, $u \leqslant_{k-1} u'$ and $v \leqslant_{k-1} v'$.

Proof. Assume first that $w \leqslant_k w'$. We have to prove that the two items in the lemma hold. For the first item, observe that by definition, $\mathcal{C} \subseteq Pol_k(\mathcal{C})$. Therefore, $w \leqslant_k w' \Rightarrow w \leqslant_\mathcal{C} w'$. We turn to the second item. Assume that $k > 0$ and consider a decomposition $w = uav$ of w. We have to find an appropriate decomposition of w'. Let $K_u = \{u' \in A^* \mid u \leqslant_k k-1u'\}$ and $K_v = \{v' \in A^* \mid v \leqslant_k k-1v'\}$. By definition K_u, K_v are upper sets for $\leqslant_k$k-1 and it follows from Lemma 2.5 that $K_u, H_v \in Pol_{k-1}(\mathcal{C})$. Hence, $K_uaK_v \in Pol_k(\mathcal{C})$ by definition. Moreover, since $w = uav \in K_uaK_v$ and $w \leqslant_k w'$, it follows that $w' \in K_uaK_v$. Therefore, we obtain $u' \in K_u$ and $v' \in K_v$ such that $w' = u'av'$. It is then immediate by definition of K_u and K_v that $u \leqslant_{k-1} u'$ and $v \leqslant_{k-1} v'$.

We turn to the other direction. Assuming that the two items in the lemma hold, we prove that $w \leqslant_k w'$. When $k = 0$, this is immediate from the first item since $Pol_0(\mathcal{C}) = \mathcal{C}$ (and therefore, $\leqslant_0$ and $\leqslant_\mathcal{C}$ are the same relation). We now assume that $k \geq 1$. Let $L \in Pol_k(\mathcal{C})$, we have to prove that $w \in L \Rightarrow w' \in L$. By definition, L is constructed by applying finitely many unions and intersections to the two following kinds of languages:

(1) Languages in $\mathcal{C}$.
(2) Languages of the form L_1aL_2 with $L_1, L_2 \in Pol_{k-1}(\mathcal{C})$.

The proof is by induction on this construction.

- When $L \in \mathcal{C}$ the implication is immediate since $w \leqslant_\mathcal{C} w'$ by the first item.
- Assume now that $L = L_1aL_2$ with $L_1, L_2 \in Pol_{k-1}(\mathcal{C})$. If $w \in L = L_1aL_2$, then it admits a decomposition $w = uav$ with $u \in L_1$ and $v \in L_2$. By the second item, we obtain $u', v' \in A^*$ such that $w' = u'av'$, $u \leqslant_{k-1} u'$ and $v \leqslant_{k-1} v'$. In particular, since $L_1, L_2 \in Pol_{k-1}(\mathcal{C})$, it follows by definition of $\leqslant_{k-1}$ that $u' \in L_1$ and $v' \in L_2$, i.e. $w' = u'av' \in L_1aL_2 = L$.
- Finally, if $L = L_1 \cup L_2$ or $L = L_1 \cap L_2$, we obtain from induction that $w \in L_1 \Rightarrow w' \in L_1$ and $w \in L_2 \Rightarrow w' \in L_2$ and therefore, $w \in L \Rightarrow w' \in L$.

This terminates the proof of Lemma 4.5. □

We now use Lemma 4.5 to present and prove two characteristic property of the preorders $\leqslant_k$ which we shall use multiple times. Let us first present the following fact which introduces a characteristic natural number of $\mathcal{C}$ that is involved in both properties.

Fact 4.6. *There exists a natural number $p \geq 1$ such that for any word $u \in A^*$ and natural numbers $m, m' \geq 1$, we have $u^{pm} \leqslant_\mathcal{C} u^{pm'}$.*

Proof. Let $\sim$ be the equivalence generated by $\leqslant_\mathcal{C}$. That is, for any $w, w' \in A^*$, $w \sim w'$ if and only if $w \leqslant_\mathcal{C} w'$ and $w' \leqslant_\mathcal{C} w$. We know from Lemmas 2.5 and 2.8 that $\leqslant_\mathcal{C}$ has finitely upper sets and is compatible with word concatenation. Therefore, $\sim$ is a congruence of finite index for word concatenation. it follows that the quotient set $A^*/{\sim}$ is a finite monoid. We let $p = \omega(A^*/{\sim})$ (the idempotent power of $A^*/{\sim}$). The fact is now immediate. □

We shall call the natural number $p \geq 1$ described in Fact 4.6 the *period of the quotienting lattice* $\mathcal{C}$. We are now ready to state the first of our two properties.

Lemma 4.7. *Let p be the period of $\mathcal{C}$ and $k \in \mathbb{N}$. Then, for any $m, m' \geq 2^{k+1} - 1$ and any word $u \in A^*$, we have $u^{pm} \leqslant_k u^{pm'}$.*

Proof. Let $m, m' \geq 2^{k+1} - 1$ and u some word, we prove that $u^{pm} \leqslant_k u^{pm'}$. This amounts to proving that the two items in Lemma 4.5 hold. The argument is an induction on k. For the first item, it suffices to prove that $u^{pm} \leqslant_{\mathcal{C}} u^{pm'}$. This is immediate by choice of p in Fact 4.6. This concludes the case $k = 0$.

We now consider the second item (which may only happen when $k \geq 1$). Consider a decomposition $u^{pm} = w_1 a w_2$. We have to find a decomposition $u^{pm'} = w_1' a w_2'$ such that $w_1 \leqslant_{k-1} w_1'$ and $w_2 \leqslant_{k-1} w_2'$. By definition, the letter a in the decomposition $u^{pm} = w_1 a w_2$ falls within some factor u^p of u^{pm}. Let us refine the decomposition to isolate this factor. We have $u^{pm} = u^{pm_1} v_1 a v_2 u^{pm_2}$ where,

- $m = m_1 + 1 + m_2$
- $v_1 a v_2 = u^p$.
- $u^{pm_1} v_1 = w_1$ and $v_2 u^{pm_2} = w_2$.

Since $m \geq 2^{k+1} - 1$ by hypothesis and $m = m_1 + 1 + m_2$, either $m_1 \geq 2^k - 1$ or $m_2 \geq 2^k - 1$ (possibly both). By symmetry, let us assume that $m_1 \geq 2^k - 1$. We use the following claim.

Claim. *There exist $m_1', m_2' \geq 1$ such that $m' = m_1' + 1 + m_2'$, $u^{pm_1} \leqslant_{k-1} u^{pm_1'}$ and $u^{pm_2} \leqslant_{k-1} u^{pm_2'}$.*

Proof. There are two cases depending on whether $m_2 \geq 2^k - 1$ or not. Assume first that, $m_2 \geq 2^k - 1$. Since $m' \geq 2^{k+1} - 1$, we may choose $m_1', m_2' \geq 2^k - 1$ such that $m' = m_1' + 1 + m_2'$. It is now immediate from induction on k that $u^{pm_1} \leqslant_{k-1} nu^{pm_1'}$ and $u^{pm_2} \leqslant_{k-1} u^{pm_2'}$. Otherwise, $m_2 < 2^k - 1$. We let $m_2' = m_2$ and $m_1' = m' - 1 - m_2'$. Clearly, $m_1' \geq 2^k - 1$ since $m' \geq 2^{k+1} - 1$. Hence, we get $u^{pm_1} \leqslant_{k-1} u^{pm_1'}$ from induction on k. Furthermore, $u^{pm_2} \leqslant_{k-1} u^{pm_2'}$ is immediate since $m_2 = m_2'$ by definition. $\square$

We may now finish the proof of Item 2. Let $m_1', m_2' \geq 1$ be as defined in the claim. We let $w_1' = u^{pm_1'} v_1$ and $w_2' = v_2 u^{pm_2'}$. Clearly, $w_1' a w_2' = u^{pm'}$ since $v_1 a v_2 = u^p$ and $m' = m_1' + 1 + m_2'$. Moreover, since $\leqslant_{k-1}$ is compatible with multiplication, we have

$$\begin{array}{lcl} w_1 = u^{pm_1} v_1 & \leqslant_{k-1} & u^{pm_1'} v_1 = w_1' \\ w_2 = v_2 u^{pm_2} & \leqslant_{k-1} & v_2 u^{pm_2'} = w_2' \end{array}$$

This terminates the proof of Item 2. $\square$

We turn to the second lemma which states a characteristic property of classes built with polynomial closure.

Lemma 4.8. *Let p be the period of $\mathcal{C}$ and $k \in \mathbb{N}$. Let $u, v \in A^*$ such that $u^p \leqslant_{\mathcal{C}} v$. Then, for any $m, m_1', m_2' \geq 2^{k+1} - 1$, we have $u^{pm} \leqslant_k u^{pm_1'} v u^{pm_2'}$.*

Proof. The proof is similar to that of Lemma 4.7. Let $k \in \mathbb{N}$, u, v satisfying $v \leqslant_{\mathcal{C}} u^p$, and $m, m_1', m_2' \geq 2^{k+1} - 1$. We prove that $u^{pm} \leqslant_k u^{pm_1'} v u^{pm_2'}$. This amounts to proving that the two items in Lemma 4.5 hold. The argument is an induction on k.

For Item 1, we prove that $u^{pm} \leqslant_{\mathcal{C}} u^{pm_1'} v u^{pm_2'}$. By hypothesis on u, v, we know that $u^p \leqslant_{\mathcal{C}} v$. Therefore, since $\leqslant_{\mathcal{C}}$ is compatible with concatenation, we get that $u^{p(m_1'+1+m_2')} \leqslant_{\mathcal{C}}$

$u^{pm'_1}vu^{pm'_2}$. Finally, we obtain by choice of p in Fact 4.6 that $u^{pm} \leqslant_{\mathcal{C}} u^{p(m'_1+1+m'_2)}$. This finishes the proof of Item 1 (and the case $k=0$) by transitivity.

We now consider the second item (which may only happen when $k \geq 1$). Consider a decomposition $u^{pm} = w_1aw_2$. We have to find a decomposition $u^{pm'_1}vu^{pm'_2} = w'_1aw'_2$ such that $w_1 \leqslant_{k-1} w'_1$ and $w_2 \leqslant_{k-1} w'_2$. By definition, the letter a in the decomposition $u^{pm} = w_1aw_2$ falls within some factor u^p of u^{pm}. Let us refine the decomposition to isolate this factor. We have $u^{pm} = u^{pm_1}v_1av_2u^{pm_2}$ where,

- $m = m_1 + 1 + m_2$
- $v_1av_2 = u^p$.
- $u^{pm_1}v_1 = w_1$ and $v_2u^{pm_2} = w_2$.

Since $m \geq 2^{k+1}-1$ by hypothesis and $m = m_1+1+m_2$, either $m_1 \geq 2^k-1$ or $m_2 \geq 2^k-1$ (possibly both). By symmetry, let us assume that $m_1 \geq 2^k-1$. We use the following claim.

Claim. *There exist $\ell'_1, \ell'_2 \in \mathbb{N}$ such that $m'_2 = \ell'_1 + 1 + \ell'_2$, $u^{pm_1} \leqslant_{k-1} u^{pm'_1}vu^{p\ell'_1}$ and $u^{pm_2} \leqslant_{k-1} u^{p\ell'_2}$.*

Proof. There are two cases depending on whether $m_2 \geq 2^k-1$ or not. Assume first that, $m_2 \geq 2^k-1$. Since $m'_2 \geq 2^{k+1}-1$, we may choose $\ell'_1, \ell'_2 \geq 2^k-1$ such that $m'_2 = \ell'_1+1+\ell'_2$. That $u^{pm_1} \leqslant_{k-1} u^{pm'_1}vu^{p\ell'_1}$ follows from induction on k. Moreover, we know that the inequality $u^{pm'_2} \leqslant_{k-1} u^{p\ell'_2}$ holds thanks to Lemma 4.7.

Otherwise, $m_2 < 2^k-1$. We let $\ell'_2 = m_2$ and $\ell'_1 = m'_2-1-\ell'_2$. Clearly, $\ell'_1 \geq 2^k-1$ since $m'_2 \geq 2^{k+1}-1$. Hence, we get $u^{pm_1} \leqslant_{k-1} u^{pm'_1}vu^{p\ell'_1}$ from induction on k. Furthermore, $u^{pm_2} \leqslant_{k-1} u^{p\ell'_2}$ is immediate since $m_2 = \ell'_2$ by definition. □

We may now finish the proof of Item 2. Let $\ell'_1, \ell'_2 \geq 1$ be as defined in the claim. We let $w'_1 = u^{pm'_1}vu^{p\ell'_1}v_1$ and $w'_2 = v_2u^{p\ell'_2}$. Clearly, $w'_1aw'_2 = u^{pm'_1}vu^{pm'_2}$ since $v_1av_2 = u^p$ and $m'_2 = \ell'_1+1+\ell'_2$. Moreover, since $\leqslant_{k-1}$ is compatible with multiplication, we have

$$\begin{array}{rcl} w_1 = u^{pm_1}v_1 & \leqslant_{k-1} & u^{pm'_1}vu^{p\ell'_1}v_1 = w'_1 \\ w_2 = v_2u^{pm_2} & \leqslant_{k-1} & v_2u^{pm'_2} = w'_2 \end{array}$$

This terminates the proof of Item 2. □

5. Framework Part 1: Covering and regularity

This is the first of two sections in which we introduce the general framework that we shall then use for formulating our covering algorithms. This framework was originally designed in [PZ16, PZ17b]. We use the more recent terminology of [PZ17b] and thus refer the reader to this paper for a more detailed presentation. In this first section, we explain how the fact that our inputs are regular languages is exploited.

When trying to decide covering (or separation) for some class $\mathcal{D}$, a key idea is that one always looks at several inputs simultaneously. Consider an input pair $(L_1, \mathbf{L}_2)$ (L_1 is a language and $\mathbf{L}_2$ a finite multiset of languages). Rather than directly deciding whether $(L_1, \mathbf{L}_2)$ is $\mathcal{D}$-coverable, our approach considers **two** multisets of languages $\mathbf{H}_1$ and $\mathbf{H}_2$ (built from L_1 and $\mathbf{L}_2$ respectively) and solves $\mathcal{D}$-covering for *all* pairs $(H_1, \mathbf{H}'_2)$ were $H \in \mathbf{H}_1$ and $\mathbf{H}'_2 \subseteq \mathbf{H}_2$. This may seem counter intuitive: we replace one input with several. Yet, the key idea is that $\mathbf{H}_1$ and $\mathbf{H}_2$ enjoy special properties which are crucially exploited by our algorithms.

Remark 5.1. *Note that this is also why our approach requires considering the covering problem even if we are only interested in separation. Even if we start with an input of the form* $(L_1, \{L_2\})$ *(the special case of separation, see Fact 2.2), we always end up working with two multisets whose sizes are larger than one.*

In this section, we present the properties that we require of our inputs and explain why we may restrict ourselves to them without loss of generality. Our algorithms are restricted to input multisets having *two* properties of different importance. The first one is common to most covering algorithms and was introduced in [PZ17b]: our inputs must be *multiplicative*. The second one is specific to the classes that we consider (*i.e.*, $Pol(\mathcal{C})$ and $PBPol(\mathcal{C})$ for some finite quotienting Boolean algebra $\mathcal{C}$): our inputs must be $\mathcal{C}$*-compatible*.

5.1. **Multiplicative multisets.** Let $\mathbf{L}$ be a finite multiset of languages. We say that $\mathbf{L}$ is *multiplicative* when $\bigcup_{L \in \mathbf{L}} L = A^*$ and $\mathbf{L}$ is equipped with a semigroup multiplication "$\odot$" (we use this notation to avoid confusion with language concatenation) satisfying the following properties:

(1) For all $L_1, L_2 \in \mathbf{L}$, we have $L_1 L_2 \subseteq L_1 \odot L_2$.
(2) For all $L \in \mathbf{L}$ and any word $w \in L$, if w can be decomposed as $w = u_1 u_2$, then there exist $L_1, L_2 \in \mathbf{L}$ such that $u_1 \in L_1$, $u_2 \in L_2$ and $L = L_1 \odot L_2$.
(3) Any element $E \in \mathbf{L}$ such that $\varepsilon \in E$ is idempotent: $E \odot E = E$.

When working with multiplicative multisets, we implicitly assume that we have the multiplication "$\odot$" in hand. The notion is designed to capture the following typical example.

Example 5.2. *Let* $\mathcal{A} = (Q, I, F, \delta)$ *be a nondeterministic finite automaton. We define,*

$$\mathbf{L}_{\mathcal{A}} = \{L_{q,r} \mid (q, r) \in Q^2\} \cup \{A^*\}$$

Here, $L_{q,r}$ *denotes the language of words labeling a run between the states* q *and* r *in* $\mathcal{A}$*. Let us point out that working with multisets is important here: when* $(q, r) \neq (q', r')$*, we need* $L_{q,r}$ *and* $L_{q',r'}$ *to be two distinct elements even if they are the same language. The multiplication "*$\odot$*" on* $\mathbf{L}_{\mathcal{A}}$ *is defined as follows. The language* A^* *is a zero:* $A^* \odot L = L \odot A^* = A^*$ *for any* $L \in \mathbf{L}_{\mathcal{A}}$*. For* $q, r, s, t \in Q$*,* $L_{q,r} \odot L_{s,t} = A^*$ *when* $r \neq s$ *and* $L_{q,r} \odot L_{s,t} = L_{q,t}$ *if* $r = s$*.*

5.2. $\mathcal{C}$**-compatible Multisets.** We further restrict our algorithms to $\mathcal{C}$-compatible inputs. This notion depends on an arbitrary finite quotienting Boolean algebra $\mathcal{C}$ which we fix for the definition. Contrary to the multiplicative property, this notion is specific to classes of the form $Pol(\mathcal{C})$ and $PBPol(\mathcal{C})$ (it is also used in [PZ17c] for $BPol(\mathcal{C})$).

Recall that since $\mathcal{C}$ is a finite Boolean algebra, one may associate a canonical equivalence $\sim_{\mathcal{C}}$ over A^* to $\mathcal{C}$: two words are equivalent when they belong to the same languages in $\mathcal{C}$. Given $w, w' \in A^*$,

$$w \sim_{\mathcal{C}} w' \quad \text{if and only if} \quad \forall L \in \mathcal{C},\ w \in L \Leftrightarrow w' \in L.$$

Given a word w, we denote by $[w]_{\mathcal{C}}$ its $\sim_{\mathcal{C}}$-equivalence class. Recall that by Lemma 2.7, $\sim_{\mathcal{C}}$ has finite index and the languages in $\mathcal{C}$ are exactly the unions of $\sim_{\mathcal{C}}$-classes. Moreover, since $\mathcal{C}$ is quotienting, we know from Lemma 2.8 that $\sim_{\mathcal{C}}$ is a congruence for word concatenation. It follows that the quotient set $A^*/{\sim_{\mathcal{C}}}$ is a finite monoid and the map $w \mapsto [w]_{\mathcal{C}}$ is a morphism from A^* to $A^*/{\sim_{\mathcal{C}}}$ (this means that $A^*/{\sim_{\mathcal{C}}}$ is a multiplicative set).

.

We are now ready to define $\mathcal{C}$-compatibility. Let $\mathbf{L}$ be any finite multiset of languages. We say that $\mathbf{L}$ *is $\mathcal{C}$-compatible* when all $L \in \mathbf{L}$ satisfy the two following conditions:

(1) L is non-empty.
(2) L is included in some equivalence class of $\sim_{\mathcal{C}}$.

Observe that if $\mathbf{L}$ is $\mathcal{C}$-compatible, then given any $L \in \mathbf{L}$, one may define the *$\mathcal{C}$-type* of L as the unique equivalence class containing L, denoted by $[L]_{\mathcal{C}}$. That is, $[L]_{\mathcal{C}} = [w]_{\mathcal{C}}$ for any word $w \in L$. Notice that in particular, $L \subseteq [L]_{\mathcal{C}} \in \mathcal{C}$.

5.3. **Reduction to multiplicative and $\mathcal{C}$-compatible multisets.** It remains to explain why we may restrict ourselves to $\mathcal{C}$-compatible multiplicative input multisets. The reason stems from the notion of *extension*, introduced in [PZ17b].

Consider two finite multisets of languages $\mathbf{H}$ and $\mathbf{L}$, we say that $\mathbf{H}$ *extends* $\mathbf{L}$ when any language in $\mathbf{L}$ is a union of languages in $\mathbf{H}$: for any $L \in \mathbf{L}$, there exists $\mathbf{H}' \subseteq \mathbf{H}$ with $L = \bigcup_{H \in \mathbf{H}'} H$. Moreover, we shall say that a subset $\mathbf{H}'$ of $\mathbf{H}$ is $\mathbf{L}$*-exhaustive* when for any $L \in \mathbf{L}$, there exists $H' \in \mathbf{H}'$ such that $H' \subseteq L$.

That we may restrict ourselves to input multisets which are both $\mathcal{C}$-compatible and multiplicative is an immediate consequence of the two following lemmas.

Lemma 5.3. *Let $\mathcal{D}$ be a lattice and consider a pair $(L_1, \mathbf{L}_2)$ where L_1 is a language and $\mathbf{L}_2$ a finite multiset of languages. Moreover, let $\mathbf{H}_1$ and $\mathbf{H}_2$ be two multisets which extend $\{L_1\}$ and $\mathbf{L}_2$ respectively. The two following properties are equivalent:*

(1) $(L_1, \mathbf{L}_2)$ is $\mathcal{D}$-coverable.
(2) For any $H_1 \in \mathbf{H}_1$ and any $\mathbf{L}_2$-exhaustive $\mathbf{H}'_2 \subseteq \mathbf{H}_2$, the pair $(H_1, \mathbf{H}'_2)$ is $\mathcal{D}$-coverable.

Lemma 5.4. *Given as input a finite multiset of regular languages $\mathbf{L}$, one may construct a finite multiplicative and $\mathcal{C}$-compatible multiset $\mathbf{H}$ extending $\mathbf{L}$ from nondeterministic finite automata recognizing the languages in $\mathbf{L}$.*

We refer the reader to [PZ17b] for the proof of these two lemmas. Let us point out that both arguments are fairly simple (in fact, half of the construction involved in the proof of Lemma 5.4 is already described in Example 5.2). Let us explain how these results are used for restricting covering to multiplicative and $\mathcal{C}$-compatible inputs.

Consider some class $\mathcal{D}$ and assume that we are trying to decide whether an input $(L_1, \mathbf{L}_2)$ is $\mathcal{D}$-coverable. Our approach works in two steps as follows:

(1) Build multiplicative and $\mathcal{C}$-compatible multisets $\mathbf{H}_1$ and $\mathbf{H}_2$ extending $\{L_1\}$ and $\mathbf{L}_2$ respectively (this is achieved with Lemma 5.4).
(2) Decide $\mathcal{D}$-coverability for *all* pairs in $(H_1, \mathbf{H}'_2) \in \mathbf{H}_1 \times 2^{\mathbf{H}_2}$. By Lemma 5.3 this information is enough to decide whether the original pair $(L_1, \mathbf{L}_2)$ is $\mathcal{D}$-coverable.

Therefore, we shall now focus on the following problem. For the two classes $\mathcal{D}$ that we investigate (i.e. $Pol(\mathcal{C})$ and $PBPol(\mathcal{C})$), we want an algorithm which takes as input two multiplicative and $\mathcal{C}$-compatible multisets $\mathbf{L}_1, \mathbf{L}_2$ and computes *all* $\mathcal{D}$-coverable pairs in $\mathbf{L}_1 \times 2^{\mathbf{L}_2}$. We explain how to approach this new problem in the next section.

6. Framework Part 2: Optimal covers

We present the second part of the general framework that we use for formulating our covering algorithms. Again, the notions that we introduce here were originally designed in [PZ16, PZ17b]. Note that we only introduce what we need in order to present our results. We refer the reader to [PZ17b] (which uses more recent terminology than [PZ16]) for details and background on these notions.

As we explained in the previous section, for the two classes $\mathcal{D}$ that we consider ($Pol(\mathcal{C})$ and $PBPol(\mathcal{C})$), our objective is getting an algorithm which takes as input two multiplicative and $\mathcal{C}$-compatible multisets $\mathbf{L}_1, \mathbf{L}_2$ and computes *all* $\mathcal{D}$-coverable pairs in $\mathbf{L}_1 \times 2^{\mathbf{L}_2}$. Here, we reformulate this problem using a more abstract framework. Specifically, we replace the second multiset $\mathbf{L}_2$ by a new algebraic object that we call *multiplicative rating map*.

Intuitively, multiplicative rating maps measure the quality of $\mathcal{D}$-covers. Assume that we have a multiplicative rating map ρ in hand. Moreover, consider some language L. We may use ρ to rank the existing $\mathcal{D}$-covers of L. This leads to the definition of a "**best**' $\mathcal{D}$-cover of L for ρ (we speak of optimal $\mathcal{D}$-cover). The key idea is we are able to reformulate our main problem with these notions. Instead of deciding which pairs in $\mathbf{L}_1 \times 2^{\mathbf{L}_2}$ are $\mathcal{D}$-coverable, we now want to compute optimal $\mathcal{D}$-covers of all $L_1 \in \mathbf{L}_1$ for a multiplicative rating map $\rho_{\mathbf{L}_2}$ that we may build from $\mathbf{L}_2$. We have two motivations for relying on this approach,

(1) It yields elegant formulations for covering algorithms. Beyond the two that we present in the paper, we refer the reader to [PZ17b] for more examples.
(2) More importantly, recall that our main goal in the paper is $PBPol(\mathcal{C})$-covering. However, achieving this objective requires a result for $Pol(\mathcal{C})$ which is stronger than the decidability of $Pol(\mathcal{C})$-covering (this explains why we also reprove the result of [PZ17c] for $Pol(\mathcal{C})$-covering along the way). The framework presented here is exactly what we need in order to precisely formulate this stronger result.

We start by defining multiplicative rating maps. Then, we explain how they are used to measure the quality of a cover and define optimal covers. Finally, we connect these notions to the covering problem. Let us point out that most of the statements presented here are without proof, we refer the reader to [PZ17b] for these proofs.

6.1. Multiplicative rating maps.

In order to present multiplicative rating maps, we need to introduce a new algebraic structure: *hemirings*. A hemiring is a set R equipped with two binary operations called addition ("+") and multiplication ("·") respectively. Moreover, the following axioms have to be satisfied:

- R is a commutative monoid for addition. The neutral element is denoted by 0_R.
- R is a semigroup for multiplication.
- Multiplication distributes over addition: for all $r, s, t \in R$ we have,

$$\begin{array}{rcl} t \cdot (r+s) & = & (t \cdot r) + (t \cdot s) \\ (r+s) \cdot t & = & (r \cdot t) + (s \cdot t) \end{array}$$

- 0_R is a zero for the multiplication: for any $r \in R$:

$$0_R \cdot r = r \cdot 0_R = 0_R$$

Remark 6.1. *Hemirings are a generalization of the more standard notion of semiring which additionally asks for the multiplication to have a neutral element.*

Finally, a hemiring R is *idempotent* when all elements are idempotents for addition: for all $r \in R$, we have $r + r = r$. In the paper, we only work with idempotent hemirings. Observe that when R is such a hemiring, one may define a *canonical partial order* on R. Given $r, s \in R$, we shall write $r \leq s$ when $s + r = s$. One may verify that this is indeed a partial order (the fact that R is idempotent is required here) which is compatible with addition and multiplication.

Example 6.2. *The most simple example of idempotent hemiring (which is crucial here) is the set 2^{A^*} of all languages over A. Indeed, it suffices to use union as the addition (the neutral element is $\emptyset$) and language concatenation as the multiplication. Observe that the canonical partial order is inclusion ($L \subseteq H$ if and only if $H \cup L = H$). In fact, 2^{A^*} has even more structure: it is a semiring ($\{\varepsilon\}$ is neutral for multiplication).*

We may now define multiplicative rating maps. We call *multiplicative rating map* a hemiring morphism $\rho : 2^{A^*} \to R$ into a *finite* idempotent hemiring R. Specifically, ρ has to satisfy the following axioms:

(1) $\rho(\emptyset) = 0_R$.
(2) For any $K_1, K_2 \subseteq A^*$, $\rho(K_1 \cup K_2) = \rho(K_1) + \rho(K_2)$.
(3) For any $K_1, K_2 \subseteq A^*$, $\rho(K_1K_2) = \rho(K_1) \cdot \rho(K_2)$.

For the sake of improved readability, when applying a multiplicative rating map ρ to a singleton language $K = \{w\}$, we shall write $\rho(w)$ for $\rho(\{w\})$. Note that ρ is compatible with the canonical orders on 2^{A^*} and R (this is true for any morphism of idempotent hemirings).

Fact 6.3. *Consider a multiplicative rating map $\rho : 2^{A^*} \to R$. Then, for any two languages $K_1, K_2 \subseteq A^*$, the following property holds, $K_1 \subseteq K_2 \Rightarrow \rho(K_1) \leq \rho(K_2)$.*

We now define a special class of multiplicative rating maps which is crucial: it is used for the connection with covering. Consider a multiplicative rating map $\rho : 2^{A^*} \to R$. We say that ρ is **nice** if for any language $K \subseteq A^*$, there exist words $w_1, \dots, w_n \in K$ such that,

$$\rho(K) = \rho(w_1) + \cdots + \rho(w_n)$$

Remark 6.4. *Nice multiplicative rating maps are finitely representable (which is not the case for arbitrary ones). Indeed, when $\rho : 2^{A^*} \to R$ is nice, it is characterized by its restriction to A^* (i.e. to singletons languages) which corresponds to a morphism between the monoid A^* and the multiplicative semigroup R (which is finitely representable). Thus, it will make sense to speak of algorithms which take nice multiplicative rating maps as input.*

Finally, observe that when we have a nice multiplicative rating map ρ in hand, it is possible to evaluate $\rho(K)$ when K is a regular language. Indeed, $\rho(K)$ is the sum of all elements $\rho(w)$ for $w \in K$ which is simple to evaluate (see [PZ17b] *for details).*

6.2. **Multiplicative rating map associated to a multiplicative multiset.** Recall that our intention is to use multiplicative rating maps for replacing one of the two multiplicative multisets in our input for the covering problem. We explain how to proceed here.

Consider an arbitrary multiplicative multiset of languages $\mathbf{L}$. We associate a canonical **nice** multiplicative rating map $\rho_{\mathbf{L}}$ to it. Since $\mathbf{L}$ is a multiplicative, the powerset $2^{\mathbf{L}}$ is a finite idempotent hemiring: the addition is union (hence, the neutral element is $\emptyset$ and the order is inclusion) and the multiplication is defined as follows. Given $\mathbf{L}_1, \mathbf{L}_2 \in 2^{\mathbf{L}}$, we define,

$$\mathbf{L}_1 \odot \mathbf{L}_2 = \{L_1 \odot L_2 \mid L_1 \in \mathbf{L}_1 \text{ and } L_2 \in \mathbf{L}_2\}$$

It is straightforward to verify from the axioms of multiplicative multi sets (see [PZ17b]) that this gives an idempotent hemiring structure to $2^{\mathbf{L}}$ and that the following map $\rho_{\mathbf{L}}$ is a **nice** multiplicative rating map:

$$\begin{array}{rccl}\rho_{\mathbf{L}}: & 2^{A^*} & \to & 2^{\mathbf{L}} \\ & K & \mapsto & \{L \in \mathbf{L} \mid K \cap L \neq \emptyset\}\end{array}$$

We call $\rho_{\mathbf{L}}$ the *canonical multiplicative rating map* associated to $\mathbf{L}$.

The $\mathcal{C}$-compatibile property. Our input multisets for covering are not only multiplicative but also $\mathcal{C}$-compatible for some finite quotienting Boolean algebra $\mathcal{C}$. Let us explain how this property translates on the associated canonical multiplicative rating maps.

Remark 6.5. *The definition of $\mathcal{C}$-compatible multiplicative rating maps that we use here is distinct from the one of* [PZ17b]. *We make this choice because our definition is simpler and suffices for our results. However, this is harmless: the two definitions coincide for nice multiplicative rating maps which is the only case where $\mathcal{C}$-compatibility is used here.*

Given an arbitrary multiplicative rating map $\rho : 2^{A^*} \to R$, we explain how one may associate a canonical multiplicative multiset of languages $\mathbf{H}_\rho$ to ρ. Then, we define the $\mathcal{C}$-compatible multiplicative rating maps as those having a $\mathcal{C}$-compatible canonical multiset.

Let $R_{A^*} \subseteq R$ be the set $R_{A^*} = \{\rho(w) \mid w \in A^*\}$ (i.e. R_{A^*} is the set of all elements in R which have a singleton antecedent). Since ρ is a morphism for multiplication, it follows that R_{A^*} is a semigroup for multiplication (in fact, it is a even a monoid). For any $s \in R_{A^*}$, we associate the language $H_s = \{w \in A^* \mid \rho(w) = s\}$ (note that H_s has to be nonempty by definition of R_{A^*}). The canonical multiset of languages $\mathbf{H}_\rho$ associated to ρ is as follows,

$$\mathbf{H}_\rho = \{H_s \mid s \in R_{A^*}\}$$

Clearly, $\mathbf{H}_\rho$ is in bijection with R_{A^*} and one may verify that it is multiplicative when equipped with the multiplication of R_{A^*}. Given $H_s, H_t \in \mathbf{H}_\rho$, their multiplication is defined by $H_s \odot H_t = H_{st}$.

We shall say that a multiplicative rating map ρ is $\mathcal{C}$-compatible when the associated canonical multiplicative multiset $\mathbf{H}_\rho$ is $\mathcal{C}$-compatible.

We finish by proving that when we start with some $\mathcal{C}$-compatible multiplicative multiset $\mathbf{L}$, then $\rho_{\mathbf{L}}$ is $\mathcal{C}$-compatible as well (i.e. $\mathbf{H}_{\rho_{\mathbf{L}}}$ is $\mathcal{C}$-compatible).

Fact 6.6. *Let $\mathbf{L}$ be a $\mathcal{C}$-compatible multiplicative multiset. Then the associated multiplicative rating map $\rho_{\mathbf{L}} : 2^{A^*} \to 2^{\mathbf{L}}$ is $\mathcal{C}$-compatible.*

Proof. Consider $H \in \mathbf{H}_{\rho_{\mathbf{L}}}$, we have to show that H is nonempty and included in a $\sim_{\mathcal{C}}$-class. By definition, there exists $u \in A^*$ such that $H = \{w \in A^* \mid \rho_{\mathbf{L}}(w) = \rho_{\mathbf{L}}(u)\}$. Thus, H is nonempty: it contains u. We show that $H \subseteq [u]_{\mathcal{C}}$. Let $w \in H$ and consider $L \in \mathbf{L}$ such that $w \in L$ (L exists since we have $\bigcup_{L \in \mathbf{L}} L = A^*$ by definition of multiplicative multisets). By definition of $\rho_{\mathbf{L}}$, we have $L \in \rho_{\mathbf{L}}(w) = \rho_{\mathbf{L}}(u) \in 2^{\mathbf{L}}$. Thus, w and u both belong to L and since $\mathbf{L}$ is $\mathcal{C}$-compatible, we obtain that $w \in [L]_{\mathcal{C}} = [u]_{\mathcal{C}}$. $\square$

6.3. **Imprints and optimal covers.** We may now explain how we use multiplicative rating maps to measure the quality of covers. This is based on a new notion called "imprints". Consider a multiplicative rating map $\rho : 2^{A^*} \to R$ (possibly not nice). For any finite set of languages $\mathbf{K}$, the ρ-imprint of $\mathbf{K}$ (denoted by $\mathcal{I}[\rho](\mathbf{K}) \subseteq R$) is the following set:

$$\mathcal{I}[\rho](\mathbf{K}) = \{r \in R \mid \text{there exists } K \in \mathbf{K} \text{ such that } r \leq \rho(K)\} \subseteq R$$

When using this notion, we will always have some language $L \subseteq A^*$ in hand and our objective will be to find the "best possible" cover $\mathbf{K}$ of L. Intuitively, ρ-imprints are designed for this purpose: given a candidate cover $\mathbf{K}$ of $\mathbf{L}$, we use the ρ-imprint of $\mathbf{K}$ to measure its "quality".

This leads to the notion of optimality. Assume that some arbitrary lattice $\mathcal{D}$ is fixed and consider a language L. An *optimal $\mathcal{D}$-cover of L for ρ* is a $\mathcal{D}$-cover of L which has the smallest possible ρ-imprint (with respect to inclusion). That is, $\mathbf{K}$ is an optimal $\mathcal{D}$-cover of L for ρ if and only if,

$$\mathcal{I}[\rho](\mathbf{K}) \subseteq \mathcal{I}[\rho](\mathbf{K}') \quad \text{for any } \mathcal{D}\text{-cover } \mathbf{K}' \text{ of } L$$

In general, there can be infinitely many optimal $\mathcal{D}$-covers of L for ρ. However, there always exists at least one (we need the fact that $\mathcal{D}$ is a lattice for this, see [PZ17b] for the proof).

Lemma 6.7. *Let $\mathcal{D}$ be a lattice. Then, for any multiplicative rating map $\rho : 2^{A^*} \to R$ and any language $L \subseteq A^*$, there exists an optimal $\mathcal{D}$-cover of L for ρ.*

Note that the proof of Lemma 6.7 is non-constructive. In fact, given L and $\rho : 2^{A^*} \to R$, computing an actual optimal $\mathcal{D}$-cover of L for ρ is a difficult problem in general. As seen in Theorem 6.10 below, when we work with the multiplicative rating map $\rho_{\mathbf{L}}$ associated to some multiplicative multiset $\mathbf{L}$, this solves $\mathcal{D}$-covering for all pairs $(L, \mathbf{L}')$ with $\mathbf{L}' \subseteq \mathbf{L}$. Before we can present this theorem, we need a key observation about optimal $\mathcal{D}$-covers.

By definition, all optimal $\mathcal{D}$-covers of L for ρ have the same ρ-imprint. Hence, this unique ρ-imprint is a *canonical* object for $\mathcal{D}$, L and ρ. We call it the *$\mathcal{D}$-optimal ρ-imprint on L* and we denote it by $\mathcal{I}_{\mathcal{D}}[L, \rho]$:

$$\mathcal{I}_{\mathcal{D}}[L, \rho] = \mathcal{I}[\rho](\mathbf{K}) \subseteq R \quad \text{for any optimal } \mathcal{D}\text{-cover of } L \text{ for } \rho.$$

We finish the definitions with two simple facts about optimal imprints which will be useful (their proofs are available in [PZ17b]).

Fact 6.8. *Consider a multiplicative rating map $\rho : 2^{A^*} \to R$ and a language L. Let $\mathcal{C}, \mathcal{D}$ be two lattices such that $\mathcal{C} \subseteq \mathcal{D}$. Then, $\mathcal{I}_{\mathcal{D}}[L, \rho] \subseteq \mathcal{I}_{\mathcal{C}}[L, \rho]$.*

Fact 6.9. *Consider a multiplicative rating map $\rho : 2^{A^*} \to R$ and $\mathcal{D}$ a lattice. Let H, L be two languages such that $H \subseteq L$. Then, $\mathcal{I}_{\mathcal{D}}[H, \rho] \subseteq \mathcal{I}_{\mathcal{D}}[L, \rho]$.*

6.4. **Connection with covering.** We may now connect optimal imprints to the covering problem. This is achieved in the following theorem (whose proof is available in [PZ17b])

Theorem 6.10. *Let $\mathcal{D}$ be a lattice. Consider a language $L \subseteq A^*$ and a multiplicative multiset $\mathbf{L}$. Recall that $\rho_{\mathbf{L}} : A^* \to 2^{\mathbf{L}}$ denotes the canonical nice multiplicative rating map associated to $\mathbf{L}$. Given any subset $\mathbf{L}' \subseteq \mathbf{L}$, the following properties are equivalent:*

(1) $(L, \mathbf{L}')$ *is $\mathcal{D}$-coverable.*
(2) $\mathbf{L}' \notin \mathcal{I}_{\mathcal{D}}[L, \rho_{\mathbf{L}}]$.

(3) Any optimal $\mathcal{D}$-cover $\mathbf{K}$ of L for $\rho_{\mathbf{L}}$ is a separating $\mathcal{D}$-cover of $(L, \mathbf{L}')$.

Theorem 6.10 formally connects the notions of this section to our problem. Recall that our objective was as follows. We wanted an algorithm which takes as input two $\mathcal{C}$-compatible multisets $\mathbf{L}_1, \mathbf{L}_2$ and computes *all* $\mathcal{D}$-coverable pairs in $\mathbf{L}_1 \times 2^{\mathbf{L}_2}$ (where $\mathcal{D}$ is either $Pol(\mathcal{C})$ or $PBPol(\mathcal{C})$). Using the second item in the theorem, we are able abstract $\mathbf{L}_2$ by the associated canonical multiplicative rating map $\rho_{\mathbf{L}_2}$. In other words, our problem is now to find an algorithm which takes the two following objects as input:

- A $\mathcal{C}$-compatible multiplicative multiset $\mathbf{L}$ (corresponding to $\mathbf{L}_1$ above).
- A nice $\mathcal{C}$-compatible multiplicative rating map ρ (corresponding to $\mathbf{L}_2$ above).

This algorithm has to compute all $\mathcal{D}$-optimal ρ-imprints $\mathcal{I}_{\mathcal{D}}[L, \rho]$ for $L \in \mathbf{L}$. Let us introduce a last notation which captures all these objects in a single one. Given a lattice $\mathcal{D}$, a multiplicative multiset $\mathbf{L}$ and a multiplicative rating map $\rho : 2^{A^*} \to R$, we shall write $\mathcal{P}_{\mathcal{D}}[\mathbf{L}, \rho]$ the following set,

$$\mathcal{P}_{\mathcal{D}}[\mathbf{L}, \rho] = \{(L, r) \in \mathbf{L} \times R \mid r \in \mathcal{I}_{\mathcal{D}}[L, \rho]\} \subseteq \mathbf{L} \times R$$

We call $\mathcal{P}_{\mathcal{D}}[\mathbf{L}, \rho]$ the *$\mathcal{D}$-optimal $\mathbf{L}$-pointed ρ-imprint.* Clearly, it encodes all sets $\mathcal{I}_{\mathcal{D}}[L, \rho]$ for $L \in \mathbf{L}$. We now have the following lemma.

Lemma 6.11. *Let $\mathcal{C}$ be a finite quotienting Boolean algebra and $\mathcal{D}$ a lattice. Assume that there exists an algorithm which computes $\mathcal{P}_{\mathcal{D}}[\mathbf{L}, \rho]$ from a $\mathcal{C}$-compatible multiplicative multiset $\mathbf{L}$ and a nice $\mathcal{C}$-compatible multiplicative rating map ρ. Then, $\mathcal{D}$-covering is decidable.*

Proof. As we explained in Section 5, it suffices to have an algorithm which takes as input two $\mathcal{C}$-compatible multisets $\mathbf{L}_1, \mathbf{L}_2$ and computes *all* $\mathcal{D}$-coverable pairs in $\mathbf{L}_1 \times 2^{\mathbf{L}_2}$. We have this algorithm since the desired information is encoded in $\mathcal{P}_{\mathcal{D}}[\mathbf{L}_1, \rho_{\mathbf{L}_2}]$ by Theorem 6.10. Moreover, we may compute this set by hypothesis. ☐

In view of Lemma 6.11, when we tackle $Pol(\mathcal{C})$- and $PBPol(\mathcal{C})$-covering in the next two sections, we shall do so by presenting algorithms which compute $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L}, \rho]$ and $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$ from a $\mathcal{C}$-compatible multiplicative multiset $\mathbf{L}$ and a nice $\mathcal{C}$-compatible multiplicative rating map $\rho : 2^{A^*} \to R$.

A key point is that these two algorithms will be presented as elegant characterization theorems. This is a design principle of our framework. Given an arbitrary quotienting lattice $\mathcal{D}$ (such as $Pol(\mathcal{C})$ or $PBPol(\mathcal{C})$), the idea is to characterize $\mathcal{P}_{\mathcal{D}}[\mathbf{L}, \rho] \subseteq \mathbf{L} \times R$ as the smallest subset of $\mathbf{L} \times R$ which contains basic elements and is closed under a set of operations. In turn this yields a least fixpoint algorithm for computing $\mathcal{P}_{\mathcal{D}}[\mathbf{L}, \rho]$: one starts from the set of basic elements and saturates it with the operations.

Remark 6.12. *There lies the difference between our results for $Pol(\mathcal{C})$ and $PBPol(\mathcal{C})$. In the case of $Pol(\mathcal{C})$, our characterization of $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L}, \rho]$ holds for* **any** *multiplicative rating map ρ (and in particular for those which are not nice). Of course, considering nice $\mathcal{C}$-compatible multiplicative rating maps suffices to get an algorithm for $Pol(\mathcal{C})$-covering (this is Lemma 6.11). However, having a characterization of $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L}, \rho]$ which holds for* **all** *multiplicative rating maps (even if it is non-effective in general) is crucial for handling $PBPol(\mathcal{C})$-covering as well. On the other hand, our characterization of $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$ only holds for nice multiplicative rating maps.*

6.5. **Properties of optimal pointed imprints.** We finish with a few results about optimal pointed imprints (i.e. the objects that we now want to compute) that we shall use in our proofs. First, we present a key lemma which states generic properties of optimal *pointed* imprints (they are satisfied by $\mathcal{P}_{\mathcal{D}}[\mathbf{L},\rho]$ for any quotienting lattice $\mathcal{D}$, any multiplicative multiset $\mathbf{L}$ and any multiplicative rating map ρ). They are involved in our characterizations of both $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ and $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L},\rho]$.

Given any multiplicative rating map $\rho : 2^{A^*} \to R$ and any multiplicative multiset $\mathbf{L}$, we shall write $\mathcal{P}_{triv}[\mathbf{L},\rho]$ for the following set,

$$\mathcal{P}_{triv}[\mathbf{L},\rho] = \{(L,r) \in \mathbf{L} \times R \mid \text{there exists } w \in L \text{ such that } r \leq \rho(w)\}$$

Let us point out that given a multiplicative multiset $\mathbf{L}$ and a **nice** multiplicative rating map ρ as input, one may compute $\mathcal{P}_{triv}[\mathbf{L},\rho]$ (see [PZ17b]). The following lemma is proved in [PZ17b].

Lemma 6.13. *Let $\mathcal{D}$ be a quotienting lattice of regular languages, $\mathbf{L}$ a multiplicative multiset and $\rho : 2^{A^*} \to R$ a multiplicative rating map. Then, $\mathcal{P}_{\mathcal{D}}[\mathbf{L},\rho] \subseteq \mathbf{L} \times R$ contains the set $\mathcal{P}_{triv}[\mathbf{L},\rho]$ and satisfies the two following closure properties:*

1. **Downset.** *For any $(L,r) \in \mathcal{P}_{\mathcal{D}}[\mathbf{L},\rho]$ and $r' \leq r$, $(L,r') \in \mathcal{P}_{\mathcal{D}}[\mathbf{L},\rho]$.*
2. **Multiplication.** *For any $(L_1,r_1),(L_2,r_2) \in \mathcal{P}_{\mathcal{D}}[\mathbf{L},\rho]$, $(L_1 \odot L_2, r_1r_2) \in \mathcal{P}_{\mathcal{D}}[\mathbf{L},\rho]$.*

Finally, we present alternate definitions of optimal pointed imprints which will be useful in proofs. The first one is restricted to classes which are finite and the second to classes which admit a stratification. Given a preorder relation $\leqslant$ defined on A^*, a word $w \in A^*$ and a language $K \subseteq A^*$, we write $w \leqslant K$ to denote the fact that $w \leqslant u$ for all $u \in K$.

Lemma 6.14. *Assume that $\mathcal{D}$ is a* **finite** *lattice. Let $\rho : 2^{A^*} \to R$ be a multiplicative rating map and $\mathbf{L}$ a multiplicative multiset. Given $(L,r) \in \mathbf{L} \times R$, the following are equivalent:*

1. *$(L,r) \in \mathcal{P}_{\mathcal{D}}[\mathbf{L},\rho]$.*
2. *There exist $w \in L$ and $K \subseteq A^*$ such that $w \leqslant_{\mathcal{D}} K$ and $r \leq \rho(K)$.*

Lemma 6.15. *Let $\mathcal{D}$ be a quotienting lattice admitting a stratification. For $k \in \mathbb{N}$, we write $\leqslant_k$ the canonical preorder of the stratum $\mathcal{D}_k$. Let $\rho : 2^{A^*} \to R$ be a multiplicative rating map and $\mathbf{L}$ a multiplicative multiset. Given $(L,r) \in \mathbf{L} \times R$, the following are equivalent:*

1. *$(L,r) \in \mathcal{P}_{\mathcal{D}}[\mathbf{L},\rho]$.*
2. *For all $k \in \mathbb{N}$, there exist $w \in L$ and $K \subseteq A^*$ such that $w \leqslant_k K$ and $r \leq \rho(K)$.*

7. Covering for $Pol(\mathcal{C})$

In this section, we present the first of our two covering algorithms. Given an arbitrary finite quotienting Boolean algebra $\mathcal{C}$ (which we fix for the section), we show that $Pol(\mathcal{C})$-covering is decidable. As announced, we use the framework introduced in Sections 5 and 6: we present an effective characterization of $Pol(\mathcal{C})$-optimal pointed imprints. More precisely, our main theorem is a description of the set $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ for any multiplicative $\mathcal{C}$-compatible multiset of languages $\mathbf{L}$ and any multiplicative rating map ρ (nice or not). Furthermore, this description yields an algorithm for computing $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ when ρ is nice.

Remark 7.1. *Naturally, the algorithm only applies to nice multiplicative rating maps. It does not even make sense to speak of an algorithm which takes arbitrary multiplicative rating maps as input: we cannot represent them finitely in general. However, the characterization* *itself applies to* **any** *multiplicative rating map. While this is useless for $Pol(\mathcal{C})$-covering, having this result will be crucial in Section 8 when we tackle $PBPol(\mathcal{C})$-covering.*

We start by presenting an auxiliary result about factorization forests that we shall need for proving our characterization (we choose to isolate it from the main argument as we shall reuse it later when considering $PBPol(\mathcal{C})$). We then present our characterization and devote the remainder of the section to its proof.

7.1. **Covering languages of factorization forests.** Consider a morphism $\alpha : A^* \to M$ into a finite monoid M. Recall that given $s \in M$ and $h, p \in \mathbb{N}$, $F^\alpha(s, h, p)$ (resp. $F^\alpha_I(s, h, p)$) denotes the language of all $w \in \alpha^{-1}(s)$ admitting an α-factorization forest of height at most h and idempotent height at most g (resp. whose root is an idempotent node). Moreover, consider a multiplicative rating map $\rho : 2^{A^*} \to R$. The following lemma holds:

Lemma 7.2. *Let $h, p \geq 1$ and $e \in M$ an idempotent. Consider a cover $\mathbf{U}$ of $F^\alpha(e, h-1, p-1)$ and $V \subseteq A^*$ containing $\alpha^{-1}(e)$. There exists a cover $\mathbf{K}$ of $F^\alpha_I(e, h, p)$ such that any $K \in \mathbf{K}$ is a concatenation $K = K_1 \cdots K_n$ with each K_i having one of the two following forms:*

(1) K_i is a language in $\mathbf{U}$, or,
(2) $K_i = U_1 \cdots U_m V U'_1 \cdots U'_{m'}$ where $U_1, \cdots, U_m, U'_1, \dots, U'_{m'} \in \mathbf{U}$ and there exists some **idempotent** *$f \in R$ such that $\rho(U_1 \cdots U_m) = \rho(U'_1 \cdots U'_{m'}) = f$.*

We now prove Lemma 7.2. Consider $h, p \geq 1$ and an idempotent $e \in M$. Finally, we let $\mathbf{U}$ as a cover of $F^\alpha(e, g-1, h-1)$ and $V \subseteq A^*$ such that $\alpha^{-1}(e) \subseteq V$. We construct a cover $\mathbf{K}$ of $F^\alpha_I(e, h, p)$ satisfying the properties described in the lemma.

For the proof, we fix $k = 2^{3|R|}$. Given any $n \geq 1$, we define $\mathbf{K}_n$ as the set of all languages $K \subseteq A^*$ which are of the form $K = K_1 \cdots K_\ell$ with $\ell \leq n$ where each K_i is of one of the two following kinds:

(1) $K_i \in \mathbf{U}$, or,
(2) $K_i = U_1 \cdots U_m V U'_1 \cdots U'_{m'}$ where $m, m' \leq k$ and $U_1, \cdots, U_m, U'_1, \dots, U'_{m'} \in \mathbf{U}$ satisfy $\rho(U_1 \cdots U_m) = \rho(U'_1 \cdots U'_{m'}) = f$ for some *idempotent* $f \in R$.

Observe that by definition all sets $\mathbf{K}_n$ are finite (each $K \in \mathbf{K}_n$ is a concatenation of at most $n \times (2k+1)$ languages in the finite set $\mathbf{U} \cup \{V\}$). Moreover all languages in $\mathbf{K}_n$ are of the form described in Lemma 7.2. Thus, it suffices to prove that there exists $n \geq 1$ such that $\mathbf{K}_n$ is a cover of $F^\alpha_I(e, h, p)$. This is what we do.

We start with some terminology. Observe that by definition of α-factorization forests, for any word $w \in F^\alpha_I(e, h, p)$, there exists a sequence of words $w_1, \dots, w_\ell \in F^\alpha(e, h-1, p-1)$ such that $w = w_1 \cdots w_\ell$. Given any such sequence $w_1, \dots, w_\ell \in F^\alpha(e, h-1, p-1)$, we associate a number that we call its *index* (we shall use this number as an induction parameter). For the definition, we fix an arbitrary linear order on $\mathbf{U}$, the cover of $F^\alpha(e, h-1, p-1)$. Moreover, for any $w \in F^\alpha(e, h-1, p-1)$, we write $U[w]$ for the smallest language in $\mathbf{U}$ containing w.

Consider a sequence $w_1, \dots, w_\ell \in F^\alpha(e, h-1, p-1)$ and an idempotent $f \in R$. We say that f *occurs in the sequence* $w_1, \dots, w_\ell$ when there exists $i \leq \ell$ and $q \leq k$ such that,

$$\rho(U[w_i] \cdots U[w_{i+q-1}]) = f$$

Finally, we define the index of the sequence $w_1, \dots, w_\ell$ as the number of *distinct idempotents* $f \in R$ that occur in $w_1, \dots, w_\ell$. Clearly, the index of $w_1, \dots, w_\ell$ is bounded by $|R|$ (note that this bound is independent from the length ℓ of the sequence). We shall need the following fact which is proved using the factorization forest theorem.

Fact 7.3. *Consider a sequence $w_1, \dots, w_\ell \in F^\alpha(e, h-1, p-1)$ of length $\ell \geq k = 2^{3|R|}$. Then the index of $w_1, \dots, w_\ell$ is larger than one: an idempotent $f \in R$ occurs in $w_1, \dots, w_\ell$.*

Proof. For all $i \leq k$, we let $r_i = \rho(U[w_i]) \in R$. Consider the morphism $\beta : R^* \to R$ which is defined by $\beta(r) = r$ for all $r \in R$. It is immediate from Theorem 4.1 that the word $(r_1) \cdots (r_k) \in R^*$ admits a β-factorization forest of height at most $3|R| - 1$. Since $k = 2^{3|R|}$, this forest has to contain at least one idempotent node. Thus, we have $i \leq k$ and $q \leq k$ such that $r_i \cdots r_{i+q-1}$ is an idempotent f of R. This concludes the proof. ☐

We may now come back to the proof of Lemma 7.2 and exhibit $n \geq 1$ such that $\mathbf{K}_n$ is a cover of $F^\alpha_I(e, h, p)$. We use the following lemma.

Lemma 7.4. *Let $d \in \mathbb{N}$ and consider a sequence $w_1, \dots, w_\ell \in F^\alpha(e, h-1, p-1)$ whose index is smaller than d. There exists a language $K \in \mathbf{K}_{(k+1)(d+1)}$ containing $w_1 \cdots w_\ell$.*

Lemma 7.2 is an immediate consequence of Lemma 7.4: we obtain that $\mathbf{K}_n$ is a cover of $F^\alpha_I(e, h, p)$ for $n = (k+1)(|R|+1)$. Indeed, by definition, for any $w \in F^\alpha_I(e, h, p)$ there exist $w_1, \dots, w_\ell \in F^\alpha(e, h-1, p-1)$ such that $w = w_1 \cdots w_\ell$. Since the index of $w_1, \dots, w_\ell$ is smaller than $|R|$, it follows from Lemma 7.4 that there exist $K \in \mathbf{K}_n$ which contains $w = w_1 \cdots w_\ell$.

It remains to prove Lemma 7.4. Let $d \in \mathbb{N}$ and consider a sequence $w_1, \dots, w_\ell \in F^\alpha(e, h-1, p-1)$ whose index is smaller than d. We construct $K \in \mathbf{K}_{(k+1)(d+1)}$ which contains $w_1 \cdots w_\ell$. The argument is an induction on d.

Assume first that $d = 0$. In that case, there exists no idempotent $f \in R$ occurring in $w_1, \dots, w_\ell$. It is immediate from Fact 7.3 that $\ell < k$. Thus, we may simply choose,

$$K = U[w_1] \cdots U[w_\ell]$$

Clearly $K \in \mathbf{K}_{(k+1)(d+1)}$ since $\ell < k$ and we have $w_1 \cdots w_\ell \in K$.

Assume now that $d \geq 1$: some idempotent $f \in R$ occurs in $w_1, \dots, w_\ell$. Thus, there exists $i \leq \ell$, $q \leq k$ such that,

$$\rho(U[w_i] \cdots U[w_{i+q-1}]) = f$$

Moreover, it follows from Fact 7.3, that we may choose i and q so that $i+q-1 \leq k$. Consider the sequence $w_{i+q+1}, \dots, w_\ell$. We distinguish two sub-cases depending on whether f occurs in $w_{i+q+1}, \dots, w_\ell$ as well or not.

Case 1: The idempotent f does not occur in $w_{i+q+1}, \dots, w_\ell$. In that case, the index of $w_{i+q+1}, \dots, w_\ell$ is smaller than $d-1$. Thus, induction yields a language $K' \in \mathbf{K}_{(k+1)d}$ containing $w_{i+q+1} \cdots w_\ell$. We define our new language K as follows:

$$K = U[w_1] \cdots U[w_{i+q}] \cdot K'$$

By definition, we have $w_1 \cdots w_\ell \in K$. Moreover, K is the concatenation of a language in $\mathbf{K}_{k+1}$ (i.e. $U[w_1] \cdots U[w_{i+q}]$ since $i+q-1 \leq k$) with a language of $\mathbf{K}_{(k+1)d}$ (i.e. K'). Thus, since $k+1+(k+1)d = (k+1)(d+1)$, we obtain that $K \in \mathbf{K}_{(k+1)(d+1)}$.

Case 2: The idempotent f occurs in $w_{i+q+1}, \dots, w_\ell$. By definition, this means that we get $j \geq i+q+1$ and $r \leq k$, such that,

$$\rho(U[w_j] \cdots U[w_{j+r-1}]) = f$$

We work with the largest such j. In other words, we choose j so that f does not occur in $w_{j+r},\dots,w_\ell$. Therefore, the index of $w_{j+r},\dots,w_\ell$ is smaller than $d-1$ and induction yields a language $K' \in \mathbf{K}_{(k+1)d}$ containing $w_{j+r}\cdots w_\ell$. We define our new language K as follows:

$$K = U[w_1]\cdots U[w_{i-1}]\cdot (U[w_i]\cdots U[w_{i+q-1}]\cdot V \cdot U[w_j]\cdots U[w_{j+r-1}])\cdot K'$$

Clearly, $U[w_1]\cdots U[w_{i-1}]\cdot (U[w_i]\cdots U[w_{i+q-1}]\cdot V \cdot U[w_j]\cdots U[w_{j+r-1}]) \in \mathbf{K}_{k+1}$ since $i \leq k+1$. Thus, K is the concatenation of a language in $\mathbf{K}_{k+1}$ with another language in $\mathbf{K}_{(k+1)d}$ and we get $K \in \mathbf{K}_{(k+1)(d+1)}$.

Finally, $w_1\cdots w_{i+q-1} \in U[w_1]\cdots U[w_{i+q-1}]$ and $w_j\cdots w_\ell \in U[w_j]\cdots U[w_{j+r-1}]\cdot K'$ by definition. Moreover, since $w_{i+q},\dots,w_{j-1} \in F^\alpha(e,h-1,p-1)$ and e is idempotent, we have $w_{i+q}\cdots w_{j-1} \in \alpha^{-1}(e) \subseteq V$. Altogether, this means that $w_1\cdots w_\ell \in K$ which concludes the proof.

7.2. **Characterization.** We begin with the property characterizing $Pol(\mathcal{C})$-optimal pointed imprints. Consider a multiplicative $\mathcal{C}$-compatible multiset $\mathbf{L}$ and a multiplicative rating map $\rho : 2^{A^*} \to R$ (note that there is no constraint on ρ). In particular, recall that since $\mathbf{L}$ is $\mathcal{C}$-compatible, for any $L \in \mathbf{L}$, $[L]_\mathcal{C}$ is well-defined as the unique $\sim_\mathcal{C}$-class containing L.

We say that a subset $S \subseteq \mathbf{L} \times R$ is *$Pol(\mathcal{C})$-saturated* (for ρ) when it contains $\mathcal{P}_{triv}[\mathbf{L},\rho]$ and is closed under the following operations:

(1) *Downset*: For any $(L,r) \in S$ and $r' \leq r$, we have $(L,r') \in S$.
(2) *Multiplication*: For any $(L_1,r_1),(L_2,r_2) \in S$, we have,
$$(L_1 \odot L_2, r_1r_2) \in S$$
(3) *$Pol(\mathcal{C})$-closure*: For any idempotent $(E,e) \in S$, we have,
$$(E, e\cdot\rho([E]_\mathcal{C})\cdot e) \in S$$

We may now state the main theorem of the section: the $Pol(\mathcal{C})$-optimal $\mathbf{L}$-pointed ρ-imprint $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ is the smallest $Pol(\mathcal{C})$-saturated subset of $\mathbf{L}\times R$ (for inclusion).

Theorem 7.5 (Characterization of $Pol(\mathcal{C})$-optimal imprints)**.** *Let $\mathbf{L}$ be a $\mathcal{C}$-compatible multiplicative multiset of languages and $\rho : 2^{A^*} \to R$ a multiplicative rating map. Then, $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ is the smallest $Pol(\mathcal{C})$-saturated subset of $\mathbf{L}\times R$.*

When ρ is nice, Theorem 7.5 yields an algorithm for computing $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ from $\mathbf{L}$ and ρ. Indeed, one may compute the smallest $Pol(\mathcal{C})$-saturated subset of $\mathbf{L}\times R$ with a least fixpoint procedure: one starts from $\mathcal{P}_{triv}[\mathbf{L},\rho]$ and saturate this set with the three operations in the definition (it is clear that all may be implemented). Combined with Lemma 6.11, this yields the desired corollary.

Corollary 7.6. *Let $\mathcal{C}$ be a finite quotienting Boolean algebra. The $Pol(\mathcal{C})$-covering problem is decidable.*

It remains to prove Theorem 7.5. The remainder of the section is devoted to this argument. Let $\mathbf{L}$ as a multiplicative $\mathcal{C}$-compatible multiset and $\rho : 2^{A^*} \to R$ as a multiplicative rating map. We have to show that $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ is the smallest $Pol(\mathcal{C})$-saturated subset of $\mathbf{L}\times R$.

We separate the proof in two main steps. Intuitively, they correspond respectively to soundness and completeness of the least fixpoint procedure obtained from the theorem. First, we show that $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ is $Pol(\mathcal{C})$-saturated. This corresponds to soundness (the least fixpoint procedure only compute elements in $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$). Then we show that $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ is included in any $Pol(\mathcal{C})$-saturated subset $S \subseteq \mathbf{L} \times R$. This corresponds to completeness (the least fixpoint procedure computes all elements in $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$)

7.3. **Soundness.** We show that $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ is $Pol(\mathcal{C})$-saturated. Since $Pol(\mathcal{C})$ is a quotienting lattice of regular languages, we already know from Lemma 6.13 that $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ contains $\mathcal{P}_{triv}[\mathbf{L},\rho]$ and is closed under downset and multiplication. Thus, we may focus on proving that $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ satisfies $Pol(\mathcal{C})$-closure. Consider an idempotent $(E,e) \in \mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$. We have to show that,

$$(E, e \cdot \rho([E]_{\mathcal{C}}) \cdot e) \in \mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$$

The proof is based on Lemma 6.15 and the generic stratification of $Pol(\mathcal{C})$ defined in Section 4 (which we may use since $\mathcal{C}$ is a finite quotienting Boolean algebra). Recall that the strata are denoted by $Pol_k(\mathcal{C})$ and the associated canonical preorders by $\leqslant_k$. The key ingredient is Lemma 4.8 which states a property of the relation $\leqslant_k$ upon which $Pol(\mathcal{C})$-closure is based.

By Lemma 6.15, we have to show that for all $k \in \mathbb{N}$, there exist $w \in E$ and a language K such that $w \leqslant_k K$ (i.e. $w \leqslant_k u$ for all $u \in K$) and $e \cdot \rho([E]_{\mathcal{C}}) \cdot e \leq \rho(K)$. Thus, we fix some $k \in \mathbb{N}$ and exhibit the appropriate w and K.

Since we have $(E,e) \in \mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ by hypothesis, we may use the opposite direction of Lemma 6.15 to get $v \in E$ and a language H such that $v \leqslant_k H$ and $e \leq \rho(H)$. Let $p \in \mathbb{N}$ be the period of $\mathcal{C}$ (see Fact 4.6) and $\ell = p2^{k+1}$. We define,

$$w = v^\ell$$

Clearly, $w \in E$ since $v \in E$ and $E \in \mathbf{L}$ is idempotent (see Item 1 in the definition of multiplicative multisets). Thus, we now need to exhibit a language K such that $w \leqslant_k K$ and $e \cdot \rho([E]_{\mathcal{C}}) \cdot e \leq \rho(K)$. Recall that we have H such that $v \leqslant_k H$ and $e \leq \rho(H)$. We define,

$$K = H^\ell \cdot [E]_{\mathcal{C}} \cdot H^\ell$$

We finish by proving that $w \leqslant_k K$ and $e \cdot \rho([E]_{\mathcal{C}}) \cdot e \leq \rho(K)$. We start with the second property which is simpler. By definition of K and since ρ is a multiplicative rating map,

$$(\rho(H))^\ell \cdot \rho([E]_{\mathcal{C}}) \cdot (\rho(H))^\ell = \rho(K)$$

Moreover, since $e \leq \rho(H)$ by hypothesis and e is idempotent, we obtain,

$$e \cdot \rho([E]_{\mathcal{C}}) \cdot e \leq \rho(K)$$

It now remains to prove that $w \leqslant_k K$. Let $u \in K$, we show that $w \leqslant_k u$. This is where we use Lemma 4.8. By definition, we have $K = H^\ell \cdot [E]_{\mathcal{C}} \cdot H^\ell$ and by hypothesis, we have $v \leqslant_k H$. Thus, since $u \in K$ and $\leqslant_k$ is compatible with concatenation (see Lemma 2.8), we obtain that there exists some $x \in [E]_{\mathcal{C}}$ such that,

$$v^\ell \cdot x \cdot v^\ell \leqslant_k u$$

Therefore, it now suffices to show that $w \leqslant_k v^\ell \cdot x \cdot v^\ell$. It will then be immediate from transitivity that $w \leqslant_k u$ as desired. Recall that p denotes the period of $\mathcal{C}$ and $\ell = p2^{k+1}$. Since $v \in E$ and E is an idempotent of $\mathbf{L}$, we have $v^p \in E \subseteq [E]_{\mathcal{C}}$. Thus, x and v^p belong to the same $\sim_{\mathcal{C}}$-class: $[E]_{\mathcal{C}}$. We get $v^p \sim_{\mathcal{C}} x$ and it is now immediate from Lemma 4.8 that

$$v^\ell \leqslant_k v^\ell \cdot x \cdot v^\ell$$

Since $w = v^\ell$ by definition this is exactly the desired property which concludes the proof.

7.4. **Completeness.** We turn to completeness. Recall that we have a multiplicative $\mathcal{C}$-compatible multiset $\mathbf{L}$ and a multiplicative rating map $\rho : 2^{A^*} \to R$. Consider $S \subseteq \mathbf{L} \times R$ which is $Pol(\mathcal{C})$-saturated. Our objective is to prove that $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L}, \rho] \subseteq S$.

We present a generic construction which builds $Pol(\mathcal{C})$-covers $\mathbf{K}_L$ for all $L \in \mathbf{L}$ such that for any $K \in \mathbf{K}_L$, we have $(L, \rho(K)) \in S$. By definition of $Pol(\mathcal{C})$-optimal ρ-imprints, this will imply that,

$$\mathcal{I}_{Pol(\mathcal{C})}[L, \rho] \subseteq \mathcal{I}[\rho](\mathbf{K}_L) \subseteq \{r \mid (L, r) \in S\}$$

Thus, by definition of $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L}, \rho]$, we shall obtain the desired inclusion: $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L}, \rho] \subseteq S$ (recall that $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L}, \rho]$ is by definition the set $\{(L, r) \mid r \in \mathcal{I}_{Pol(\mathcal{C})}[L, \rho]\}$).

Remark 7.7. *This construction yields a generic method for building optimal $Pol(\mathcal{C})$-covers for all $L \in \mathbf{L}$. Indeed, we may apply it in the special case when $S = \mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L}, \rho]$ since we already showed above that $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L}, \rho]$ is $Pol(\mathcal{C})$-saturated. In that case, we get $Pol(\mathcal{C})$-covers $\mathbf{K}_L$ for all $L \in \mathbf{L}$ such that $\mathcal{I}[\rho](\mathbf{K}_L) = \mathcal{I}_{Pol(\mathcal{C})}[L, \rho]$.*

The construction is based on Simon's factorization forest theorem which we presented in Section 4. Using this theorem requires a monoid morphism $\alpha : A^* \to M$. We start by defining α. Recall that since $\mathbf{L}$ is multiplicative, the powerset $2^{\mathbf{L}}$ is a semigroup: given $\mathbf{L}_1, \mathbf{L}_2 \in 2^{\mathbf{L}}$, their multiplication is,

$$\mathbf{L}_1 \odot \mathbf{L}_2 = \{L_1 \odot L_2 \mid L_1 \in \mathbf{L}_1 \text{ and } L_2 \in \mathbf{L}_2\}$$

Consider the map $\alpha : A^* \to 2^{\mathbf{L}}$ defined by $\alpha(w) = \{L \in \mathbf{L} \mid w \in L\}$. Since $\mathbf{L}$ is multiplicative, one may verify that α is a semigroup morphism (α is the restriction to A^* of the canonical multiplicative rating map $\rho_{\mathbf{L}}$). Thus, the restriction of α to its image is a monoid morphism.

We may now start the construction. Recall that for all $\mathbf{T} \in 2^{\mathbf{L}}$ and $h, p \in \mathbb{N}$, we write $F^\alpha(\mathbf{T}, h, p)$ the language of all words in $\alpha^{-1}(\mathbf{T})$ which admit an α-factorization forest of height at most h and idempotent height at most p. We use the following proposition.

Proposition 7.8. *Let $\mathbf{T} \in 2^{\mathbf{L}}$ and $h, p \in \mathbb{N}$. There exists a $Pol(\mathcal{C})$-cover $\mathbf{K}$ of $F^\alpha(\mathbf{T}, h, p)$ such that for any $L \in \mathbf{T}$ and any $K \in \mathbf{K}$, we have $(L, \rho(K)) \in S$.*

Before proving the proposition, we use it to finish the proof of Theorem 7.5. Consider a language $L \in \mathbf{L}$. Our objective is to build a $Pol(\mathcal{C})$-cover $\mathbf{K}_L$ of L such that,

$$(L, \rho(K)) \in S \quad \text{for all } K \in \mathbf{K}_L$$

Let $h = p = 3|2^{\mathbf{L}}| - 1$. We obtain from Simon's factorization forest theorem (i.e. Theorem 4.1) that for any $\mathbf{T} \in 2^{\mathbf{L}}$, we have $\alpha^{-1}(\mathbf{T}) = F^\alpha(\mathbf{T}, h, p)$. Thus, given any $\mathbf{T} \in 2^{\mathbf{L}}$, Proposition 7.8 yields a $Pol(\mathcal{C})$-cover $\mathbf{K}_{\mathbf{T}}$ of $\alpha^{-1}(\mathbf{T})$ such that for any $K \in \mathbf{K}_{\mathbf{T}}$ and $L \in \mathbf{T}$, we have $(L, \rho(K)) \in S$. Observe that by definition of α, we have,

$$L = \bigcup_{\{\mathbf{T} \in 2^{\mathbf{L}} \mid L \in \mathbf{T}\}} \alpha^{-1}(\mathbf{T})$$

Therefore, it suffices to define $\mathbf{K}_L$ as the union of all sets $\mathbf{K}_{\mathbf{T}}$ such that $L \in \mathbf{T}$. It is immediate by definition that $\mathbf{K}_L$ is a $Pol(\mathcal{C})$-cover of L and that $(L, \rho(K)) \in S$ for all $K \in \mathbf{K}_L$.

It remains to prove Proposition 7.8. Let $h,p \in \mathbb{N}$ and $\mathbf{T} \in 2^{\mathbf{L}}$. Our objective is to build a $Pol(\mathcal{C})$-cover $\mathbf{K}$ of $F^{\alpha}(\mathbf{T},h,p)$ which satisfies the following property:

$$\text{For all } L \in \mathbf{T} \text{ and all } K \in \mathbf{K}, \quad (L,\rho(K)) \in S \tag{7.1}$$

Observe that if $\alpha^{-1}(\mathbf{T}) = \emptyset$, this is easily achieved: it suffices to define $\mathbf{K} = \emptyset$. Thus, we shall assume from now on that $\alpha^{-1}(\mathbf{T}) \neq \emptyset$. The construction is achieved by induction on the height h. We start by considering the base case: $h = 0$.

Remark 7.9. *The proof is actually independent the idempotent height p: this parameter is only useful for the $PBPol(\mathcal{C})$ proof which we present in the next section. We are forced to make it apparent here because we intend to use Lemma 7.2 whose statement is designed to accommodate both $Pol(\mathcal{C})$ and $PBPol(\mathcal{C})$.*

7.4.1. *Base case: Leaves.* Assume that $h = 0$. It follows that all words in $F^{\alpha}(\mathbf{T},0,p)$ are either empty or made of a single letter $a \in A$. We distinguish two cases depending on whether $\alpha(\varepsilon) = \mathbf{T}$. If $\alpha(\varepsilon) \neq \mathbf{T}$, we define,

$$\mathbf{K} = \{[\varepsilon]_{\mathcal{C}}a[\varepsilon]_{\mathcal{C}} \mid a \in A \text{ and } \alpha(a) = \mathbf{T}\}$$

Otherwise, $\alpha(\varepsilon) = \mathbf{T}$ and we define,

$$\mathbf{K} = \{[\varepsilon]_{\mathcal{C}}a[\varepsilon]_{\mathcal{C}} \mid a \in A \text{ and } \alpha(a) = \mathbf{T}\} \cup \{[\varepsilon]_{\mathcal{C}}\}$$

Clearly, all languages in $\mathbf{K}$ belong to $Pol(\mathcal{C})$ (they are marked concatenations of the language $[\varepsilon]_{\mathcal{C}} \in \mathcal{C}$ with itself). Moreover, it is also immediate that $\mathbf{K}$ is a cover of $F^{\alpha}(\mathbf{T},0,p)$. Indeed, a word $w \in F^{\alpha}(\mathbf{T},0,p)$ is either empty or a single letter and its image is $\mathbf{T}$ under α. Thus, if $w = \varepsilon$, we have $w \in [\varepsilon]_{\mathcal{C}}$ and if $w = a \in A$, we have $w \in [\varepsilon]_{\mathcal{C}}a[\varepsilon]_{\mathcal{C}}$.

It remains to show that $\mathbf{K}$ satisfies (7.1). Given $L \in \mathbf{T}$ and $K \in \mathbf{K}$, we have to show that $(L,\rho(K)) \in S$. By definition of $\mathbf{K}$ this is an immediate consequence of the following lemma.

Lemma 7.10. *Let $L \in \mathbf{L}$ and $a \in A$ the following properties hold:*

(1) If $\varepsilon \in L$, then $(L,\rho([\varepsilon]_{\mathcal{C}})) \in S$.
(2) If $a \in L$, then $(L,\rho([\varepsilon]_{\mathcal{C}}a[\varepsilon]_{\mathcal{C}})) \in S$.

Proof. We begin with the first item: when $\varepsilon \in L$, we have $(L,\rho([\varepsilon]_{\mathcal{C}})) \in S$. Since $\varepsilon \in L$, we have $(L,\rho(\varepsilon)) \in \mathcal{P}_{triv}[\mathbf{L},\rho]$ by definition. Therefore, since S is $Pol(\mathcal{C})$-saturated, we have,

$$(L,\rho(\varepsilon)) \in S$$

By Item 3 in the definition of multiplicative multisets $L = L \odot L$ since $\varepsilon \in L$. Moreover, $\rho(\varepsilon) = \rho(\varepsilon) \cdot \rho(\varepsilon)$ since ρ is a multiplicative rating map. Altogether, this means that $(L,\rho(\varepsilon))$ is an idempotent and since S is $Pol(\mathcal{C})$-saturated, we get from $Pol(\mathcal{C})$-closure that,

$$(L,\rho([L]_{\mathcal{C}})) = (L,\rho(\varepsilon) \cdot \rho([L]_{\mathcal{C}}) \cdot \rho(\varepsilon)) \in S$$

Finally, since $\varepsilon \in L$, we have $[L]_{\mathcal{C}} = [\varepsilon]_{\mathcal{C}}$ and we conclude that $(L,\rho([\varepsilon]_{\mathcal{C}})) \in S$ as desired.

It remains to prove the second item. Given $a \in A$ such that $a \in L$, we show that $(L,\rho([\varepsilon]_{\mathcal{C}}a[\varepsilon]_{\mathcal{C}})) \in S$. Observe that a can be decomposed as $a = \varepsilon a \varepsilon$. Hence, since $a \in L$, it follows from Item 2 in the definition of multiplicative sets that we have $L_1, L_a, L_2 \in \mathbf{L}$ such that $\varepsilon \in L_1$, $\varepsilon \in L_2$, $a \in L_a$ and $L_1 \odot L_a \odot L_2 = L$. Since $\varepsilon \in L_1, L_2$, we may reuse the above argument to obtain,

$$(L_1, \rho([\varepsilon]_{\mathcal{C}})) \in S$$
$$(L_2, \rho([\varepsilon]_{\mathcal{C}})) \in S$$

Moreover, since $a \in L_a$, we have $(L_a, \rho(a)) \in \mathcal{P}_{triv}[\mathbf{L}, \rho]$ by definition which means that $(L_a, \rho(a)) \in S$ since S is $Pol(\mathcal{C})$-saturated. Using again the fact that S is $Pol(\mathcal{C})$-saturated (specifically closure under multiplication), we obtain the following,

$$(L, \rho([\varepsilon]_{\mathcal{C}} a [\varepsilon]_{\mathcal{C}})) = (L_1 \odot L_a \odot L_2, \rho([\varepsilon]_{\mathcal{C}}) \cdot \rho(a) \cdot \rho([\varepsilon]_{\mathcal{C}})) \in S$$

This concludes the proof of Lemma 7.10. □

7.4.2. *Inductive case.* We now assume that $h \geq 1$. Recall that our objective is to build a $Pol(\mathcal{C})$-cover $\mathbf{K}$ of $F^\alpha(\mathbf{T}, h, p)$ which satisfies (7.1). We decompose $F^\alpha(\mathbf{T}, h, p)$ as the union of three languages that we cover independently.

Recall that $F^\alpha_B(\mathbf{T}, h, p)$ (resp. $F^\alpha_I(\mathbf{T}, h, p)$) denotes the language of all words in $\alpha^{-1}(\mathbf{T})$ admitting an α-factorization forest of height of at most h, of idempotent height at most p and whose root is a *binary node* (resp. *idempotent node*). The construction is based on the two following lemmas.

Lemma 7.11. *There exists a $Pol(\mathcal{C})$-cover $\mathbf{K}_B$ of $F^\alpha_B(\mathbf{T}, h, p)$ such that for all $K \in \mathbf{K}_B$ and $L \in \mathbf{T}$, we have $(L, \rho(K)) \in S$.*

Lemma 7.12. *There exists a $Pol(\mathcal{C})$-cover $\mathbf{K}_I$ of $F^\alpha_I(\mathbf{T}, h, p)$ such that for all $K \in \mathbf{K}_I$ and $L \in \mathbf{T}$, we have $(L, \rho(K)) \in S$.*

Before we show these two results, let us use them to finish the inductive case. Let $\mathbf{K}_B$ and $\mathbf{K}_I$ be as defined in the two above lemmas. Since we assumed that $h \geq 1$ it is immediate that $F^\alpha(\mathbf{T}, h, p)$ is equal to the following union:

$$F^\alpha(\mathbf{T}, h, p) = F^\alpha_B(\mathbf{T}, h, p) \cup F^\alpha_I(\mathbf{T}, h, p) \cup F^\alpha(\mathbf{T}, h-1, p)$$

Moreover, we obtain from induction on h that we have a $Pol(\mathcal{C})$-cover $\mathbf{K}'$ of $F^\alpha(\mathbf{T}, h-1, p)$ such that for all $K \in \mathbf{K}$ and any $L \in \mathbf{T}$, we have $(L, \rho(K)) \in S$. Thus, it suffices to define $\mathbf{K} = \mathbf{K}_B \cup \mathbf{K}_I \cup \mathbf{K}'$. By definition, we know that $\mathbf{K}$ is a $Pol(\mathcal{C})$-cover of $F^\alpha(\mathbf{T}, h, p)$ which satisfies (7.1) as desired. It remains to prove the two lemmas.

Proof of Lemma 7.11. Using induction we obtain that for all $\mathbf{H} \in 2^{\mathbf{L}}$, there exists a $Pol(\mathcal{C})$-cover $\mathbf{U}_{\mathbf{H}}$ of $F^\alpha(\mathbf{H}, h-1, p)$ such that for any $U \in \mathbf{U}_{\mathbf{H}}$ and any $L \in \mathbf{H}$, we have $(L, \rho(U)) \in S$. We use these new $Pol(\mathcal{C})$-covers $\mathbf{U}_{\mathbf{H}}$ to build the desired $\mathbf{K}_B$. We define,

$$\mathbf{K}_B = \{K_1K_2 \mid \text{there exist } \mathbf{H}_1, \mathbf{H}_2 \in 2^{\mathbf{L}} \text{ s.t. } \mathbf{T} = \mathbf{H}_1 \odot \mathbf{H}_2,\ K_1 \in \mathbf{U}_{\mathbf{H}_1} \text{ and } K_2 \in \mathbf{U}_{\mathbf{H}_2}\}$$

We need to verify that $\mathbf{K}_B$ satisfies the desired properties. We start by proving that it is a $Pol(\mathcal{C})$-cover of $F^\alpha_B(\mathbf{T}, h, p)$. Clearly, all languages in $\mathbf{K}_B$ belong to $Pol(\mathcal{C})$ (they are all concatenations of two languages in $Pol(\mathcal{C})$). Hence, it suffices to verify that $\mathbf{K}_B$ is a cover of $F^\alpha_B(\mathbf{T}, h, p)$. Let $w \in F^\alpha_B(\mathbf{T}, h, p)$, we exhibit $K \in \mathbf{K}_B$ such that $w \in K$. By definition, w is the root label of some α-factorization forest of height at most h, of idempotent height at most p and whose root is a binary node. This exactly says that w admits a decomposition $w = w_1w_2$ with $w_1 \in F^\alpha(\alpha(w_1), h-1, p)$, $w_2 \in F^\alpha(\alpha(w_2), h-1, p)$. Since $\mathbf{U}_{\alpha(w_1)}$ and $\mathbf{U}_{\alpha(w_2)}$ are covers of $F^\alpha(\alpha(w_1), h-1, p)$ and $F^\alpha(\alpha(w_2), h-1, p)$ respectively, there exist

$K_1 \in \mathbf{U}_{\alpha(w_1)}$ and $K_2 \in \mathbf{U}_{\alpha(w_2)}$ such that $w_1 \in K_1$ and $w_2 \in K_2$. Thus, $w = w_1w_2 \in K_1K_2$ which is an element of $\mathbf{K}_B$ by definition since $\alpha(w_1) \odot \alpha(w_2) = \alpha(w) = \mathbf{T}$.

It remains to verify that for any $K \in \mathbf{K}_B$ and $L \in \mathbf{T}$, we have $(L, \rho(K)) \in S$. By construction, $K = K_1K_2$ where $K_1 \in \mathbf{U}_{\mathbf{H}_1}$ and $K_2 \in \mathbf{U}_{\mathbf{H}_2}$ for $\mathbf{H}_1, \mathbf{H}_2 \in 2^{\mathbf{L}}$ such that $\mathbf{T} = \mathbf{H}_1 \odot \mathbf{H}_2$. Therefore, since $L \in \mathbf{T}$, we get $L_1 \in \mathbf{H}_1$ and $L_2 \in \mathbf{H}_2$ such that $L = L_1 \odot L_2$. Moreover, we know that $(L_1, \rho(K_1)) \in S$ and $(L_2, \rho(K_2)) \in S$ by definition of $\mathbf{U}_{\mathbf{H}_1}$ and $\mathbf{U}_{\mathbf{H}_2}$. Therefore, since S is $Pol(\mathcal{C})$-saturated, we obtain from closure under multiplication that $(L, \rho(K)) = (L_1 \odot L_2, \rho(K_1K_2)) \in S$.

Proof of Lemma 7.12. Recall that our objective here is to construct a $Pol(\mathcal{C})$-cover $\mathbf{K}_I$ of $F_I^\alpha(\mathbf{T}, h, p)$ such that for all $K \in \mathbf{K}_I$ and $L \in \mathbf{T}$, we have $(L, \rho(K)) \in S$. Observe that we may assume without loss of generality that $\mathbf{T}$ is an idempotent of $2^{\mathbf{L}}$. Indeed, otherwise we have $F_I^\alpha(\mathbf{T}, h, p) = \emptyset$ and we may simply choose $\mathbf{K}_I = \emptyset$. We shall write $\mathbf{T} = \mathbf{E} \in 2^{\mathbf{L}}$ to underline the fact that $\mathbf{T}$ is idempotent: we have to cover $F_I^\alpha(\mathbf{E}, h, p)$. This is where we use Lemma 7.2. Applying it requires a cover $\mathbf{U}$ of $F^\alpha(\mathbf{E}, h-1, p-1)$ and a language V which contains $\alpha^{-1}(\mathbf{E})$. Let us first define these objects.

We build $\mathbf{U}$ by induction. More precisely, it is immediate from induction on the height h that there exists a $Pol(\mathcal{C})$-cover $\mathbf{U}$ of $F^\alpha(\mathbf{E}, h-1, p-1)$ such that for any $U \in \mathbf{U}$ and any $L \in \mathbf{E}$, we have $(L, \rho(U)) \in S$. We now use the following fact to define the language V.

Fact 7.13. *There exists a $\sim_\mathcal{C}$ class V such that $[L]_\mathcal{C} = V$ for all $L \in \mathbf{E}$ and $\alpha^{-1}(\mathbf{E}) \subseteq V$.*

Proof. By hypothesis, $\alpha^{-1}(\mathbf{E}) \neq \emptyset$. Thus, we may choose some arbitrary word $w \in \alpha^{-1}(\mathbf{E})$. Let $V = [w]_\mathcal{C}$. We verify that V satisfies the desired properties. By definition of α, $\mathbf{E} = \{L \in \mathbf{L} \mid w \in L\}$. Thus, all $L \in \mathbf{E}$ must satisfy $[L]_\mathcal{C} = V$ since they contain $w \in V$. Finally, observe that $\mathbf{E}$ is nonempty (since $\mathbf{L}$ is multiplicative, we have $\bigcup_{L \in \mathbf{L}} = A^*$ and at least one $L \in \mathbf{L}$ contains w). Thus, we may choose some $L \in \mathbf{E}$. By definition all words $v \in \alpha^{-1}(\mathbf{E})$ belong to L and therefore to $[L]_\mathcal{C} = V$: we get that $\alpha^{-1}(\mathbf{E}) \subseteq V$. $\square$

We now have everything we need for applying Lemma 7.2. We obtain a cover $\mathbf{K}_I$ of $F_I^\alpha(\mathbf{E}, h, p)$ such that any $K \in \mathbf{K}_I$ is a concatenation $K = K_1 \cdots K_n$ where each K_i is of one of the two following kinds:

(1) K_i is a language in $\mathbf{U}$, or,
(2) $K_i = U_1 \cdots U_m V U'_1 \cdots U'_{m'}$ where $U_1, \dots, U_m, U'_1, \dots, U'_{m'} \in \mathbf{U}$ and there exists an *idempotent* $f \in R$ such that $\rho(U_1 \cdots U_m) = \rho(U'_1 \cdots U'_{m'}) = f$.

Clearly, any $K \in \mathbf{K}_I$ belongs to $Pol(\mathcal{C})$: it is a concatenation of languages in $Pol(\mathcal{C})$ by definition ($V \in \mathcal{C} \subseteq Pol(\mathcal{C})$ since it is $\sim_\mathcal{C}$-class). Therefore, $\mathbf{K}_I$ is a $Pol(\mathcal{C})$-cover of $F_I^\alpha(\mathbf{E}, h, p)$. It remains to prove that for any $K \in \mathbf{K}_I$ and $L \in \mathbf{E}$, we have $(L, \rho(K)) \in S$. We need the following fact which is proved using $Pol(\mathcal{C})$-closure.

Fact 7.14. *Consider an idempotent $f \in R$ such that $f = \rho(U_1 \cdots U_m)$ with $U_1, \dots, U_m \in \mathbf{U}$. For any $L \in \mathbf{E}$, we have $(L, f \cdot \rho(V) \cdot f) \in S$.*

Let us first use Fact 7.14 to finish the main proof. Consider $K \in \mathbf{K}_I$ and $L \in \mathbf{E}$, we show that $(L, \rho(K)) \in S$. By definition $K = K_1 \cdots K_n$ where all languages K_i are as described in the two items above. Since $\mathbf{E}$ is idempotent, we have $\mathbf{E} = (\mathbf{E})^n$ and since $L \in \mathbf{E}$, we get $L_1, \dots, L_n \in \mathbf{E}$ such that $L = L_1 \odot \cdots \odot L_n$. Observe that for all $i \leq n$, we have $(L_i, \rho(K_i)) \in S$. If K_i is as described in the first item ($K_i \in \mathbf{U}$), this is by definition of

$\mathbf{U}$. Otherwise, when K_i is as described in the second item, this is by Fact 7.14. Therefore, since S is $Pol(\mathcal{C})$-saturated, we obtain from closure under multiplication that,

$$(L, \rho(K)) = (L_1 \odot \cdots \odot L_n, \rho(K_1) \cdots \rho(K_n)) \in S$$

This concludes the argument for Lemma 7.12. It remains to prove Fact 7.14. We finish with this proof.

Consider an idempotent $f \in R$ such that $f = \rho(U_1 \cdots U_m)$ for some $U_1, \dots, U_m \in \mathbf{U}$ and let $L \in \mathbf{E}$. we show that $(L, f \cdot \rho(V) \cdot f) \in S$. Since $\mathbf{E}$ is an idempotent of $2^{\mathbf{L}}$, we have $\mathbf{E} = (\mathbf{E})^{|\mathbf{L}|+1}$. Thus, since $L \in \mathbf{E}$, we have $L_1, \dots, L_{|\mathbf{L}|+1} \in \mathbf{E}$ such that, $L = L_1 \odot \cdots \odot L_{|\mathbf{L}|+1}$. One may then use a pumping argument to obtain $L', L'', E \in \mathbf{E}$ such that E is an idempotent of $\mathbf{L}$ and $L = L' \odot E \odot L''$. By definition of $\mathbf{U}$, we know that for all $i \leq m$, we have $(L', \rho(U_i)) \in S$, $(L'', \rho(U_i)) \in S$ and $(E, \rho(U_i)) \in S$. Since S is $Pol(\mathcal{C})$-saturated and E is idempotent, we obtain from closure under multiplication that,

$$\begin{array}{lcl} (E, f) & = & ((E)^m, \rho(U_1 \cdots U_m)) \in S \\ (L' \odot E, f) & = & (L' \odot (E)^{m-1}, \rho(U_1 \cdots U_m)) \in S \\ (E \odot L'', f) & = & ((E)^{m-1} \odot L'', \rho(U_1 \cdots U_m)) \in S \end{array}$$

Moreover, since (E, f) is idempotent, the first equality and $Pol(\mathcal{C})$-closure yield that,

$$(E, f \cdot \rho([E]_{\mathcal{C}}) \cdot f) \in S$$

Finally, we have $[E]_{\mathcal{C}} = V$ (since $E \in \mathbf{E}$, see Fact 7.13). Thus, we get $(E, f \cdot \rho(V) \cdot f) \in S$. It then suffices to use closure under multiplication one last time to obtain,

$$(L, f \cdot \rho(V) \cdot f) = (L' \odot E \odot L'', f \cdot f \cdot \rho(V) \cdot f \cdot f) \in S$$

This concludes the proof of Fact 7.14.

8. Covering for $PBPol(\mathcal{C})$

We now turn to our main result: $PBPol(\mathcal{C})$-covering is decidable for any finite quotienting Boolean algebra $\mathcal{C}$. Again, our approach is based on optimal imprints: we present an effective characterization of $PBPol(\mathcal{C})$-optimal pointed imprints. For the sake of avoiding clutter, we shall assume that $\mathcal{C}$ is fixed for the section.

This characterization is more involved than the one we already obtained for $Pol(\mathcal{C})$. First, it applies to a more restricted class of multiplicative rating maps. Specifically, we present a characterization of $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$ which holds when $\mathbf{L}$ is a $\mathcal{C}$-compatible multiplicative multiset and ρ is a *nice $\mathcal{C}$-compatible* multiplicative rating map. In other words, we need ρ to be nice and $\mathcal{C}$-compatible (which was not the case for $Pol(\mathcal{C})$).

Remark 8.1. *Of course, having a characterization restricted to this special case is enough to obtain the desired $PBPol(\mathcal{C})$-covering algorithm by Lemma 6.11. However, the fact that we now require ρ to be nice (which was not the case for $Pol(\mathcal{C})$) is significant. This explains why the arguments of this paper do not extend to higher levels in concatenation hierarchies. The proof of our characterization for $PBPol(\mathcal{C})$ relies heavily on the fact that we have a characterization for $Pol(\mathcal{C})$ which holds for* **all** *multiplicative rating maps.*

A second point is that our characterization of $PBPol(\mathcal{C})$-optimal pointed imprints actually involves two distinct objects:

- As desired, it describes $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$, the $PBPol(\mathcal{C})$-optimal $\mathbf{L}$-pointed ρ-imprint.

- However, it also describes a second object: $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \tau]$, the $Pol(\mathcal{C})$-optimal $\mathbf{H}_\rho$-pointed τ-imprint where $\mathbf{H}_\rho$ is the canonical multiset associated to ρ, and τ is an auxiliary multiplicative rating map built from $\mathbf{L}$ and ρ.

The key idea here is that our descriptions of $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$ and $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \tau]$ are mutually dependent. Reformulated from an algorithmic point of view, this means that we get a least fixpoint procedure which computes $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$ and $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \tau]$ simultaneously.

Remark 8.2. *The presence of this second object $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \tau]$ explains why we shall need to reuse our characterization of $Pol(\mathcal{C})$-optimal pointed imprints (i.e. Theorem 7.5). In particular, the multiplicative rating map τ will* **not** *be nice. Hence, it will be important that Theorem 7.5 holds for all multiplicative rating maps.*

The section is organized as follows. First, we explain how the auxiliary multiplicative rating map τ is defined from $\mathbf{L}$ and ρ. Then, we present our characterization of $PBPol(\mathcal{C})$-optimal pointed imprints. Finally, we concentrate on proving our characterization (note that we postpone the difficult direction of this proof to the next section).

8.1. **$\mathcal{D}$-derived multiplicative rating maps.** Let $\mathcal{D}$ be a quotienting lattice which is closed under concatenation (here, we shall use the case when $\mathcal{D} = PBPol(\mathcal{C})$). Given any multiplicative multiset $\mathbf{L}$ and any multiplicative rating map $\rho : 2^{A^*} \to R$. We associate a new multiplicative rating map $\tau : 2^{A^*} \to 2^{\mathbf{L} \times R}$ called the *$\mathcal{D}$-derived multiplicative rating map associated to $\mathbf{L}$ and ρ.* Let us first explain why the set $2^{\mathbf{L} \times R}$ is a hemiring.

Since $2^{\mathbf{L} \times R}$ is a set of subsets, it is an idempotent commutative monoid for union. Thus, we simply use union as our addition (in particular, the neutral element is $\emptyset$ and the order is inclusion). It remains to define our multiplication. Given $T_1, T_2 \in 2^{\mathbf{L} \times R}$, we define,

$$T_1 \cdot T_2 = \{(L_1 \odot L_2, r) \mid \text{there exists } (L_1, r_1) \in T_1 \text{ and } (L_2, r_2) \in T_2 \text{ such that } r \leq r_1 r_2\}$$

One may verify that this is indeed a semigroup multiplication which distributes over the addition (i.e. union) and that $\emptyset$ (the neutral element for union) is a zero. It follows that $2^{\mathbf{L} \times R}$ is an idempotent hemiring.

Remark 8.3. *Our multiplication is not the most immediate one: $T_1 \cdot T_2$ is* **not** *the set of all multiplications between elements of T_1 and T_2. It contains more elements: we make sure that $T_1 \cdot T_2$ is closed under downset (if $(L, r) \in T_1 \cdot T_2$ and $r' \leq r$, then $(L, r') \in T_1 \cdot T_2$). We shall need this for proving that $\tau : 2^{A^*} \to 2^{\mathbf{L} \times R}$ is a multiplicative rating map.*

We are now ready to define our new multiplicative rating map $\tau : 2^{A^*} \to 2^{\mathbf{L} \times R}$. We use the following definition:

$$\begin{array}{rccl} \tau : & 2^{A^*} & \to & 2^{\mathbf{L} \times R} \\ & K & \mapsto & \{(L, r) \in \mathbf{L} \times R \mid r \in \mathcal{I}_{\mathcal{D}}[K \cap L, \rho]\} \end{array}$$

As announced, we call τ the *$\mathcal{D}$-derived multiplicative rating map associated to $\mathbf{L}$ and ρ.* Let us verify that τ is indeed a multiplicative rating map.

Lemma 8.4. *The map τ is a multiplicative rating map.*

Proof. We have to show that τ is a hemiring morphism. Let us first consider addition (which is union for both 2^{A^*} and $2^{\mathbf{L} \times R}$). Clearly $\tau(\emptyset) = \emptyset$. Indeed,

$$\tau(\emptyset) = \{(L, r) \in \mathbf{L} \times R \mid r \in \mathcal{I}_{\mathcal{D}}[\emptyset, \rho]\} = \{(L, r) \in \mathbf{L} \times R \mid r \in \emptyset\} = \emptyset$$

We now show that for any $K_1, K_2 \subseteq A^*$, we have $\tau(K_1 \cup K_2) = \tau(K_1) \cup \tau(K_2)$. We start with the right to left inclusion. Let $(L,r) \in \tau(K_1) \cup \tau(K_2)$ and by symmetry, assume that $(L,r) \in \tau(K_1)$. It follows that $r \in \mathcal{I}_\mathcal{D}[K_1 \cap L, \rho]$. Since $K_1 \cap L \subseteq (K_1 \cup K_2) \cap L$, it then follows from Fact 6.9 that $r \in \mathcal{I}_\mathcal{D}[(K_1 \cup K_2) \cap L, \rho]$ which exactly says that $(L,r) \in \tau(K_1 \cup K_2)$. It remains to treat the left to right inclusion. Let $(L,r) \in \tau(K_1 \cup K_2)$. By definition, $r \in \mathcal{I}_\mathcal{D}[(K_1 \cup K_2) \cap L, \rho]$. Consider optimal $\mathcal{D}$-covers (for ρ) $\mathbf{U}_1, \mathbf{U}_2$ of $(K_1 \cap L)$ and $(K_2 \cap L)$ respectively. By definition, $\mathbf{U}_1 \cup \mathbf{U}_2$ is a $\mathcal{D}$-cover of $(K_1 \cup K_2) \cap L$. Thus, we have $\mathcal{I}_\mathcal{D}[(K_1 \cup K_2) \cap L, \rho] \subseteq \mathcal{I}[\rho](\mathbf{U}_1 \cup \mathbf{U}_2)$ and we obtain $r \in \mathcal{I}[\rho](\mathbf{U}_1 \cup \mathbf{U}_2)$. Therefore, there exists $U \in \mathbf{U}_1 \cup \mathbf{U}_2$ such that $r \leq \rho(U)$. By symmetry assume that $U \in \mathbf{U}_1$, we show that $(L,r) \in \tau(K_1)$ (when $U \in \mathbf{U}_2$, one may show that $r \in \tau(K_2)$). By definition, we have $r \in \mathcal{I}[\rho](\mathbf{U}_1)$ and since $\mathbf{U}_1$ is an optimal $\mathcal{D}$-cover of $K_1 \cap L$, we know that $\mathcal{I}[\rho](\mathbf{U}_1) = \mathcal{I}_\mathcal{D}[K_1 \cap L, \rho]$. Thus, $r \in \mathcal{I}_\mathcal{D}[K_1 \cap L, \rho]$ which exactly means that $(L,r) \in \tau(K_1)$.

This concludes the proof for addition. We turn to multiplication. We show that $\tau(K_1K_2) = \tau(K_1) \cdot \tau(K_2)$. We start with the inclusion $\tau(K_1) \cdot \tau(K_2) \subseteq \tau(K_1K_2)$. Let $(L,r) \in \tau(K_1) \cdot \tau(K_2)$. Thus, we have $(L_1, r_1) \in \tau(K_1)$ and $(L_2, r_2) \in \tau(K_2)$ such that $L = L_1 \odot L_2$ and $r \leq r_1r_2$. By definition of τ, we have $r_1 \in \mathcal{I}_\mathcal{D}[K_1 \cap L_1, \rho]$ and $r_2 \in \mathcal{I}_\mathcal{D}[K_2 \cap L_2, \rho]$. Since $\mathcal{D}$ is a quotienting lattice, we get from a result of [PZ17b], that,

$$r_1r_2 \in \mathcal{I}_\mathcal{D}[(K_1 \cap L_1) \cdot (K_2 \cap L_2), \rho]$$

Observe that $(K_1 \cap L_1) \cdot (K_2 \cap L_2) \subseteq K_1K_2 \cap L_1L_2$. Moreover, by Item 1 in the definition of multiplicative sets, we have $L_1L_2 \subseteq L_1 \odot L_2 = L$. Altogether, this means that we have $(K_1 \cap L_1) \cdot (K_2 \cap L_2) \subseteq K_1K_2 \cap L$. Therefore, we then obtain from Fact 6.9 that $r_1r_2 \in \mathcal{I}_\mathcal{D}[K_1K_2 \cap L, \rho]$. Thus, we also have $r \in \mathcal{I}_\mathcal{D}[K_1K_2 \cap L, \rho]$ by definition of imprints since $r \leq r_1r_2$. It follows that $(L,r) \in \tau(K_1K_2)$ by definition of τ.

We finish with the converse inclusion. Let $(L,r) \in \tau(K_1K_2)$. By definition, we have $r \in \mathcal{I}_\mathcal{D}[K_1K_2 \cap L, \rho]$. For any language $H \in \mathbf{L}$ and $i \in \{1,2\}$, we define $\mathbf{U}_{i,H}$ as an optimal $\mathcal{D}$-cover of $K_i \cap H$. Consider the following finite set of languages $\mathbf{U}$,

$$\mathbf{U} = \{U_1U_2 \mid \text{there exists } L_1, L_2 \in \mathbf{L} \text{ s.t. } L_1 \odot L_2 = L,\ U_1 \in \mathbf{U}_{1,L_1} \text{ and } U_2 \in \mathbf{U}_{2,L_2}\}$$

We claim that $\mathbf{U}$ is a $\mathcal{D}$-cover of $K_1K_2 \cap L$. Before we prove this claim let us use it to finish the argument. Recall that $r \in \mathcal{I}_\mathcal{D}[K_1K_2 \cap L, \rho]$. Thus, since $\mathbf{U}$ is a $\mathcal{D}$-cover of $K_1K_2 \cap L$, we have $r \in \mathcal{I}[\rho](\mathbf{U})$. It follows that there exists $U \in \mathbf{U}$ such that $r \leq \rho(U)$. By definition of $\mathbf{U}$, $U = U_1U_2$ with $U_1 \in \mathbf{U}_{1,L_1}$ and $U_2 \in \mathbf{U}_{2,L_2}$ where $L_1, L_2 \in \mathbf{L}$ satisfy $L_1 \odot L_2 = L$. Let $r_1 = \rho(U_1)$ and $r_2 = \rho(U_2)$. Since $\mathbf{U}_{1,L_1}$ and $\mathbf{U}_{2,L_2}$ are optimal $\mathcal{D}$-cover of $K_1 \cap L_1$ and $K_2 \cap L_2$ respectively, we have $r_1 \in \mathcal{I}_\mathcal{D}[K_1 \cap L_1, \rho]$ and $r_2 \in \mathcal{I}_\mathcal{D}[K_2 \cap L_2, \rho]$. It follows that $(L_1, r_1) \in \tau(K_1)$ and $(L_2, r_2) \in \tau(K_2)$. Since $r \leq \rho(U) = \rho(U_1) \cdot \rho(U_2) = r_1r_2$, it follows by definition of our multiplication that $(L,r) \in \tau(K_1) \cdot \tau(K_2)$ which concludes the proof.

It remains to show that $\mathbf{U}$ is a $\mathcal{D}$-cover of $K_1K_2 \cap L$. Clearly, all languages in $\mathbf{U}$ belong to $\mathcal{D}$ by definition since $\mathcal{D}$ is closed under concatenation. Hence, it suffices to verify that $\mathbf{U}$ is a cover of $K_1K_2 \cap L$. Consider $w \in K_1K_2 \cap L$. We exhibit $U \in \mathbf{U}$ such that $w \in U$. By definition, we have $w = w_1w_2$ with $w_1 \in K_1$ and $w_2 \in K_2$. Moreover, since $w_1w_2 = w \in L$, we obtain from the definition of multiplicative multisets that there exist $L_1, L_2 \in \mathbf{L}$ such that $L_1 \odot L_2 = L$, $w_1 \in L_1$ and $w_2 \in L_2$. Altogether, this means that we have $w_1 \in K_1 \cap L_1$ and $w_2 \in K_2 \cap L_2$. Thus, we have $U_1 \in \mathbf{U}_{1,L_1}$ and $U_2 \in \mathbf{U}_{1,L_2}$ such that $w_1 \in U_1$ and $w_2 \in U_2$. We conclude that $w \in U_1U_2$ which is an element of $\mathbf{U}$ by definition. $\square$

An important point is that while the $\mathcal{D}$-derived multiplicative rating map τ associated to $\mathbf{L}$ and ρ is a multiplicative rating map, it is **not** a nice one in general. Moreover, let us point our that given a (regular) language $K \subseteq A^*$ computing its image $\tau(K)$ is a difficult task: we have to compute all $\mathcal{D}$-optimal ρ-imprints $\mathcal{I}_{\mathcal{D}}[K \cap L, \rho]$ for $L \in \mathbf{L}$.

Remark 8.5. *We shall use these definitions in the case when $\mathcal{D} = PBPol(\mathcal{C})$. We intend to consider $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \tau]$, the $Pol(\mathcal{C})$-optimal $\mathbf{H}_\rho$-pointed τ-imprint (where $\mathbf{H}_\rho$ is the canonical multiset associated to ρ but this is of little importance for the moment). Since we proved a characterization of $Pol(\mathcal{C})$-optimal pointed imprints (Theorem 7.5), we already have a lot of information on $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \tau]$. However, what we* **cannot** *do is use this characterization to directly compute $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \tau]$ from $\mathbf{L}$ and ρ. Indeed, in order to implement the third operation ($Pol(\mathcal{C})$-closure), one has to evaluate $\tau(V)$ when V is some $\sim_{\mathcal{C}}$-class. This is precisely what we cannot do.*

There is however a case when evaluating $\tau(K)$ is a simple (provided that $\mathcal{D}$ contains all finite languages which is the case for $PBPol(\mathcal{C})$ by Lemma 3.4): when K is a singleton. In that case $\tau(K)$ does not depend on $\mathcal{D}$.

Lemma 8.6. *Assume that $\mathcal{D}$ contains all finite languages and consider the $\mathcal{D}$-derived multiplicative rating map τ associated to $\mathbf{L}$ and ρ. Then for any $w \in A^*$, we have $\tau(w) = \{(L, r) \in \mathbf{L} \times R \mid w \in L \text{ and } r \leq \rho(w)\}$.*

Proof. Assume first that $(L, r) \in \tau(w)$. Then, we have $r \in \mathcal{I}_{\mathcal{D}}[\{w\} \cap L, \rho]$. It follows that $\{w\} \cap L \neq \emptyset$, i.e. $w \in L$ and we have $r \in \mathcal{I}_{\mathcal{D}}[\{w\}, \rho]$. Moreover, since $\mathcal{D}$ contains finite languages, $\mathbf{K} = \{\{w\}\}$ is a $\mathcal{D}$-cover of $\{w\}$. Hence, we have $\mathcal{I}_{\mathcal{D}}[\{w\}, \rho] \subseteq \mathcal{I}[\rho](\mathbf{K})$ which means any $r \leq \rho(w)$ by definition.

Conversely, assume that we have $(L, r) \in \mathbf{L} \times R$ such that $w \in L$ and $r \leq \rho(w)$. We show that $(L, r) \in \tau(w)$, i.e. $r \in \mathcal{I}_{\mathcal{D}}[\{w\} \cap L, \rho]$. Since $w \in L$, we have $\{w\} \cap L = \{w\}$: we have to show that $r \in \mathcal{I}_{\mathcal{D}}[\{w\}, \rho]$. This is immediate since $r \in \mathcal{I}_{triv}[\{w\}, \rho]$ by definition. ☐

Lemma 8.6 is important: if we have $\mathbf{L}$ and ρ in hand, we are able to compute the restriction of τ to A^*: the (multiplicative) monoid morphism $w \mapsto \tau(w)$. Of course, this restriction does not fully describe τ since it is not nice. However, having it is sufficient to compute the set $\mathcal{P}_{triv}[\mathbf{H}, \tau]$ for any multiplicative multiset $\mathbf{H}$. Indeed, recall that,

$$\mathcal{P}_{triv}[\mathbf{H}, \tau] = \{(H, T) \in \mathbf{H} \times 2^{\mathbf{L} \times R} \mid \text{there exists } w \in H \text{ such that } T \subseteq \tau(w)\}$$

This will be useful later (having this set is required for implementing our $PBPol(\mathcal{C})$-covering algorithm).

8.2. **Characterization.**

We now turn to our characterization of $PBPol(\mathcal{C})$-optimal pointed imprints. Let $\mathbf{L}$ be a $\mathcal{C}$-compatible multiplicative multiset of languages and $\rho : 2^{A^*} \to R$ be a $\mathcal{C}$-compatible multiplicative rating map.

Remark 8.7. *We do not ask ρ to be nice yet. The presentation of our characterization makes sense for any multiplicative rating map (but it only holds for the nice ones).*

Recall that ρ being $\mathcal{C}$-compatible means that the canonical multiplicative multiset $\mathbf{H}_\rho$ associated to ρ is $\mathcal{C}$-compatible. Moreover, recall that each $H \in \mathbf{H}_\rho$ is associated to a unique element $r \in R$: $H = \{w \in A^* \mid \rho(w) = r\}$. In other words, one may view each $H \in \mathbf{H}_\rho$ as an alias for some element of R. Finally, we write $\tau : 2^{A^*} \to 2^{\mathbf{L} \times R}$ for the $PBPol(\mathcal{C})$-derived multiplicative rating map associated to $\mathbf{L}$ and ρ.

Remark 8.8. *Since $\mathbf{L}$ and $\mathbf{H}_\rho$ are both $\mathcal{C}$-compatible, recall that given any $L \in \mathbf{L}$ (resp. $H \in \mathbf{H}_\rho$), $[L]_\mathcal{C}$ (resp. $[H]_\mathcal{C}$) is well defined as the unique $\sim_\mathcal{C}$-class containing L (resp. H).*

As announced, our characterization simultaneously describes $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho] \subseteq \mathbf{L} \times R$ and $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \tau] \subseteq \mathbf{H}_\rho \times 2^{\mathbf{L} \times R}$. We first introduce a notion of $PBPol(\mathcal{C})$-saturated object which describes the properties satisfied by these two sets. The definition applies to pairs $(S, \mathcal{T})$ where $S \subseteq \mathbf{L} \times R$ and $\mathcal{T} \subseteq \mathbf{H}_\rho \times 2^{\mathbf{L} \times R}$. We say that such a pair $(S, \mathcal{T})$ is $PBPol(\mathcal{C})$-saturated for ρ when the following conditions are satisfied:

- **Conditions on $S \subseteq \mathbf{L} \times R$.** S contains $\mathcal{P}_{triv}[\mathbf{L}, \rho]$ and is closed under the following operations:
  1. **Downset:** If $(L, r) \in S$ and $r' \leq r$, then $(L, r') \in S$.
  2. **Multiplication:** for any $(L_1, r_1), (L_2, r_2) \in S$,
  $$(L_1 \odot L_2, r_1 r_2) \in S$$
  3. $PBPol(\mathcal{C})$**-closure.** For any $(H, T) \in \mathcal{T}$, if $r \in R$ is the element such that $H = \{w \in A^* \mid \rho(w) = r\}$, given any idempotent $(E, f) \in T \subseteq \mathbf{L} \times R$,
  $$(E, f \cdot (r + \rho(\varepsilon)) \cdot f) \in S$$
- **Conditions on $\mathcal{T} \subseteq \mathbf{H}_\rho \times 2^{\mathbf{L} \times R}$.** $\mathcal{T}$ contains $\mathcal{P}_{triv}[\mathbf{H}_\rho, \tau]$ and is closed under the following operations:
  1. **Downset:** If $(H, T) \in \mathcal{T}$ and $T' \subseteq T$, then $(H, T') \in \mathcal{T}$.
  2. **Multiplication:** For any $(H_1, T_1), (H_2, T_2) \in \mathcal{T}$,
  $$(H_1 \odot H_2, T_1 T_2) \in \mathcal{T}$$
  3. **Nested closure.** For any idempotent $(E, F) \in \mathcal{T}$,
  $$(E, F \cdot T \cdot F) \in \mathcal{T} \quad \text{where } T = \{(L, r) \in S \mid [E]_\mathcal{C} = [L]_\mathcal{C}\}$$

Finally, we shall say that a set $S \subseteq \mathbf{L} \times R$ is $PBPol(\mathcal{C})$-saturated (for ρ), when there exists some $\mathcal{T} \subseteq \mathbf{H}_\rho \times 2^{\mathbf{L} \times R}$ such that the pair $(S, \mathcal{T})$ is $PBPol(\mathcal{C})$-saturated. It turns out that **when ρ is nice**, the smallest $PBPol(\mathcal{C})$-saturated subset of $\mathbf{L} \times R$ is $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$.

Theorem 8.9 (Characterization of $PBPol(\mathcal{C})$-optimal imprints)**.** *Let $\mathbf{L}$ be a $\mathcal{C}$-compatible multiplicative multiset and $\rho : 2^{A^*} \to R$ a* **nice** *$\mathcal{C}$-compatible multiplicative rating map. Then, $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$ is the smallest $PBPol(\mathcal{C})$-saturated subset of $\mathbf{L} \times R$.*

Clearly, Theorem 8.9 yields a least fixpoint algorithm for computing $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$ from $\mathbf{L}$ and $\rho : 2^{A^*} \to R$. Indeed, let $\mathbf{H}_\rho$ be the canonical multiset associated to ρ and $\tau : 2^{A^*} \to 2^{\mathbf{L} \times R}$ the $PBPol(\mathcal{C})$-derived multiplicative rating map associated to $\mathbf{L}$ and ρ. Clearly, one may compute $\mathbf{H}_\rho$ from ρ. Moreover, as we explained in Lemma 8.6, while we do not have an effective representation of τ, we are still able to compute the set $\mathcal{P}_{triv}[\mathbf{H}_\rho, \tau]$. Thus, we may use a least fixpoint procedure to compute the smallest $PBPol(\mathcal{C})$-saturated pair $(S, \mathcal{T})$: it starts from the sets $\mathcal{P}_{triv}[\mathbf{L}, \rho]$ and $\mathcal{P}_{triv}[\mathbf{H}_\rho, \tau]$ and saturates them with the six operations in the definition. By Theorem 8.9, S is the desired set $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$.

Remark 8.10. *The statement of Theorem 8.9 does not describe the smallest associated subset $\mathcal{T} \subseteq \mathbf{H}_\rho \times 2^{\mathbf{L} \times R}$ such that the pair $(\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho], \mathcal{T})$ is $PBPol(\mathcal{C})$-saturated (which we also compute). This is because, we are mainly interested in $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$: $\mathcal{T}$ is only an auxiliary object that is required for the computation. However, we shall see in the proof that $\mathcal{T}$ is exactly $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \tau]$ (as we announced above).*

Altogether, we obtain the desired corollary from Lemma 6.11: $PBPol(\mathcal{C})$-covering is decidable.

Corollary 8.11. *Let $\mathcal{C}$ be a finite quotienting Boolean algebra. The $PBPol(\mathcal{C})$-covering problem is decidable.*

We turn to the proof of Theorem 8.9. Here, we focus on the direction which corresponds to soundness of the least fixpoint procedure: we show that $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L},\rho]$ is indeed $PBPol(\mathcal{C})$-saturated. The other direction of the proof ($\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L},\rho]$ is the smallest such set) is more involved and we postpone it to the next section.

8.3. **Soundness.** Let $\mathbf{L}$ be a $\mathcal{C}$-compatible multiplicative multiset and $\rho : 2^{A^*} \to R$ a nice $\mathcal{C}$-compatible multiplicative rating map. We show that $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L},\rho]$ is $PBPol(\mathcal{C})$-saturated.

Remark 8.12. *We do not use the fact that ρ is nice here. This is only needed for the other direction of the proof.*

Let $\mathbf{H}_\rho$ be the $\mathcal{C}$-compatible canonical multiset associated to ρ and $\tau : 2^{A^*} \to 2^{\mathbf{L}\times R}$ the $PBPol(\mathcal{C})$-derived multiplicative rating map associated to $\mathbf{L}$ and ρ. We prove that,

$$\text{The pair } (\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L},\rho], \mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho,\tau]) \text{ is } PBPol(\mathcal{C})\text{-saturated}$$

Since $Pol(\mathcal{C})$ and $PBPol(\mathcal{C})$ are both quotienting lattices of regular languages, we already know from Lemma 6.13 that $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L},\rho]$ and $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{L},\rho]$ contain $\mathcal{P}_{triv}[\mathbf{L},\rho]$ and $\mathcal{P}_{triv}[\mathbf{H}_\rho,\tau]$ respectively and are both closed under downset and multiplication. Therefore, we may concentrate on $PBPol(\mathcal{C})$-closure and nested closure. We start with the latter.

8.3.1. *Nested closure.* Consider an idempotent $(E,F) \in \mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho,\tau]$, we show that,

$$(E, F\cdot T\cdot F) \in \mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho,\tau] \quad \text{where } T = \{(L,r) \in \mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L},\rho] \mid [E]_\mathcal{C} = [L]_\mathcal{C}\}$$

We use Theorem 7.5, our characterization of $Pol(\mathcal{C})$-optimal pointed imprints. It states that $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho,\tau]$ is $Pol(\mathcal{C})$-saturated (for τ). Thus, it satisfies $Pol(\mathcal{C})$-closure and we have,

$$(E, F\cdot \tau([E]_\mathcal{C})\cdot F) \in \mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho,\tau]$$

We show that $T = \tau([E]_\mathcal{C})$ which concludes the proof. By definition of τ, we know that,

$$\tau([E]_\mathcal{C}) = \{(L,r) \in \mathbf{L}\times R \mid r \in \mathcal{I}_{PBPol(\mathcal{C})}[[E]_\mathcal{C}\cap L,\rho]\}$$

Since $\mathbf{L}$ is $\mathcal{C}$-compatible, we know that for any $L \in \mathbf{L}$ the intersection $[E]_\mathcal{C}\cap L$ is either equal to L (when $[E]_\mathcal{C} = [L]_\mathcal{C}$) or empty. Thus, we have,

$$\tau([E]_\mathcal{C}) = \{(L,r) \in \mathbf{L}\times R \mid [E]_\mathcal{C} = [L]_\mathcal{C} \text{ and } r \in \mathcal{I}_{PBPol(\mathcal{C})}[L,\rho]\}$$

Since $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L},\rho] = \{(L,r) \in \mathbf{L}\times R \mid r \in \mathcal{I}_{PBPol(\mathcal{C})}[L,\rho]\}$ by definition, it follows that $\tau([E]_\mathcal{C}) = \{(L,r) \in \mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L},\rho] \mid [E]_\mathcal{C} = [L]_\mathcal{C}\} = T$ which concludes the proof.

8.3.2. *$PBPol(\mathcal{C})$-closure.* This requires more work. Consider a pair $(H,T) \in \mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \tau]$ and let $r \in R$ be the element such that $H = \{w \in A^* \mid \rho(w) = r\}$. Moreover, let $(E,f) \in T \subseteq \mathbf{L} \times R$ be an idempotent. We show that,

$$(E, f \cdot (r + \rho(\varepsilon)) \cdot f) \in \mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$$

We shall need the following simple fact. Recall that given a quotienting lattice $\mathcal{D}$, we write $\overline{\mathcal{D}}$ the quotienting lattice containing all complements of languages in $\mathcal{D}$.

Fact 8.13. *There exists a* **finite** *quotienting lattice $\mathcal{D} \subseteq Pol(\mathcal{C})$ such that $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho] = \mathcal{P}_{Pol(\overline{\mathcal{D}})}[\mathbf{L}, \rho]$.*

Proof. For all $L \in \mathbf{L}$, let $\mathbf{K}_L$ be an optimal $PBPol(\mathcal{C})$-cover of L (for ρ). Recall that we showed in Lemma 3.6 that $PBPol(\mathcal{C}) = Pol(\overline{Pol(\mathcal{C})})$. Thus, since there are finitely many languages in the sets $\mathbf{K}_L$ for $L \in \mathbf{L}$, there exists some $k \in \mathbb{N}$ such that all these languages belong to $Pol(\overline{Pol_k(\mathcal{C})})$ (recall that $Pol_k(\mathcal{C})$ denotes a stratum in our stratification of $Pol(\mathcal{C})$). It now suffices to choose $\mathcal{D} = Pol_k(\mathcal{C})$. ☐

In view of Fact 8.13, it now suffices to show that,

$$(E, f \cdot (r + \rho(\varepsilon)) \cdot f) \in \mathcal{P}_{Pol(\overline{\mathcal{D}})}[\mathbf{L}, \rho]$$

Since $\overline{\mathcal{D}}$ is a finite quotienting lattice, we have a stratification of $Pol(\overline{\mathcal{D}})$ which we introduced in Section 4. Recall that the strata are denoted by $Pol_k(\overline{\mathcal{D}})$ and the associated canonical preorder relations by $\leqslant_k$. Recall that by Lemma 6.15, it suffices to show that for any $k \in \mathbb{N}$, we have $w \in E$ and a language $K \subseteq A^*$ such that $w \leqslant_k K$ (i.e. $w \leqslant_k u$ for all $u \in K$) and $f \cdot (r + \rho(\varepsilon)) \cdot f \leq \rho(K)$. Consider $k \in \mathbb{N}$, we exhibit the appropriate w and K.

By hypothesis we have $(H,T) \in \mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \tau]$ and since $\mathcal{D} \subseteq Pol(\mathcal{C})$ it follows that $(H,T) \in \mathcal{P}_{\mathcal{D}}[\mathbf{H}_\rho, \tau]$ (see Fact 6.8). Recall that $\leqslant_{\mathcal{D}}$ denotes the canonical preorder associated to the finite quotienting lattice $\mathcal{D}$. Since, $(H,T) \in \mathcal{P}_{\mathcal{D}}[\mathbf{H}_\rho, \tau]$, it follows from Lemma 6.14 that there exist $x \in H$ and $G \subseteq A^*$ such that $x \leqslant_{\mathcal{D}} G$ and $T \subseteq \tau(G)$. Therefore, since $(E,f) \in T \subseteq \tau(G)$, we obtain by definition of τ that,

$$f \in \mathcal{I}_{PBPol(\mathcal{C})}[G \cap E, \rho]$$

Moreover, since $\mathcal{D} \subseteq Pol(\mathcal{C})$, we have $Pol(\overline{\mathcal{D}}) \subseteq PBPol(\mathcal{C})$. Therefore, it follows from Fact 6.8 that $f \in \mathcal{I}_{Pol(\overline{\mathcal{D}})}[G \cap E, \rho]$. Using Lemma 6.15, this yields $v \in G \cap E$ and $J \subseteq A^*$ such that $v \leqslant_k J$ and $f \leq \rho(J)$.

We are now ready to define the appropriate $w \in E$ and $K \subseteq A^*$. Let $p \in \mathbb{N}$ be the period of $\overline{\mathcal{D}}$ (see Fact 4.6) and $\ell = p \times 2^{k+1}$. We define,

$$w = v^\ell \quad \text{and} \quad K = J^\ell \cdot \{x, \varepsilon\} \cdot J^\ell$$

Clearly, $w \in E$ since $v \in E$ and E is an idempotent of $\mathbf{L}$ (see Item 1 in the definition of multiplicative multisets). It remains to show that $w \leqslant_k K$ and $f \cdot (r + \rho(\varepsilon)) \cdot f \leq \rho(K)$. We start with the latter. By definition and since ρ is a multiplicative rating map,

$$\rho(K) = (\rho(J))^\ell \cdot (\rho(x) + \rho(\varepsilon)) \cdot (\rho(J))^\ell$$

Since $x \in H$, we have $\rho(x) = r$ (H is the language $H = \{y \in A^* \mid \rho(y) = r\}$ by definition). Moreover, we have $f \leq \rho(J)$ by definition of J and f is idempotent. Thus, we obtain as desired that,

$$f \cdot (r + \rho(\varepsilon)) \cdot f \leq \rho(K)$$

We finish with the proof that $w \leqslant_k K$. Let $u \in K$, we show that $w \leqslant_k u$. By definition, we have $K = J^\ell \cdot \{x, \varepsilon\} \cdot J^\ell$ and $v \leqslant_k J$. Thus, since $u \in K$ and $\leqslant_k$ is compatible with concatenation (see Lemma 2.8), we obtain that one of the two following properties hold:

- $v^\ell \cdot v^\ell \leqslant_k u$, or,
- $v^\ell \cdot x \cdot v^\ell \leqslant_k u$.

Thus, by transitivity, it suffices to show that $w \leqslant_k v^\ell \cdot v^\ell$ and $w \leqslant_k v^\ell \cdot x \cdot v^\ell$. Since $w = v^\ell$ and $\ell = p2^{k+1}$ where p is the period of $\overline{\mathcal{D}}$, that $w \leqslant_k v^\ell \cdot v^\ell$ is immediate from Lemma 4.7.

It remains to show that $w \leqslant_k v^\ell \cdot x \cdot v^\ell$. Recall that $\leqslant_k$ is the canonical preorder of $Pol_k(\overline{\mathcal{D}})$. By definition, we have $v \in G$ which means that $x \leqslant_{\mathcal{D}} v$. It follows that $v \leqslant_{\overline{\mathcal{D}}} x$. Indeed, $x \leqslant_{\mathcal{D}} v$ means that for any $L \in \mathcal{D}$, we have $x \in L \Rightarrow v \in L$. The contrapositive then states that for any $L \in \mathcal{D}$, we have $v \notin L \Rightarrow x \notin L$. Finally, since the languages of $\overline{\mathcal{D}}$ are the complements of those in $\mathcal{D}$, it follows that for all $L \in \overline{\mathcal{D}}$, we have $v \in L \Rightarrow x \in L$, i.e. $v \leqslant_{\overline{\mathcal{D}}} x$. Therefore, since $w = v^\ell$ and $\ell = p2^{k+1}$ where p is the period of $\overline{\mathcal{D}}$, we obtain as desired that $w \leqslant_k v^\ell \cdot x \cdot v^\ell$ from Lemma 4.8.

9. Completeness in Theorem 8.9

This last section is devoted to proving the difficult direction of Theorem 8.9. Consider a multiplicative and $\mathcal{C}$-compatible multiset $\mathbf{L}$ and a nice $\mathcal{C}$-compatible multiplicative rating map $\rho : 2^{A^*} \to R$. Recall that we write $\mathbf{H}_\rho$ for the canonical multiset associated to ρ (which is also $\mathcal{C}$-compatible). Moreover τ is the $PBPol(\mathcal{C})$-derived multiplicative rating map associated to $\mathbf{L}$ and ρ:

$$\begin{array}{rlll} \tau : & 2^{A^*} & \to & 2^{\mathbf{L} \times R} \\ & K & \mapsto & \{(L, r) \in \mathbf{L} \times R \mid r \in \mathcal{I}_{PBPol(\mathcal{C})}[K \cap L, \rho]\} \end{array}$$

We proved in the previous section that $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$ is a $PBPol(\mathcal{C})$-saturated subset of $\mathbf{L} \times R$. It remained to show that it is the smallest such subset. Therefore, we fix some arbitrary $PBPol(\mathcal{C})$-saturated subset S of $\mathbf{L} \times R$. By definition, S is paired with a subset $\mathcal{T} \subseteq \mathbf{H}_\rho \times 2^{\mathbf{L} \times R}$ for which the pair $(S, \mathcal{T})$ is $PBPol(\mathcal{C})$-saturated. Our objective here is proving that,

$$\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho] \subseteq S$$

While the argument itself is more involved, its structure is similar to the one we used when dealing with $Pol(\mathcal{C})$ in Section 7. Specifically, we present a generic construction for building $PBPol(\mathcal{C})$-covers $\mathbf{K}_L$ for all $L \in \mathbf{L}$ such that for any $K \in \mathbf{K}_L$, we have $(L, \rho(K)) \in S$. By definition of $PBPol(\mathcal{C})$-optimal ρ-imprints, the existence of these $PBPol(\mathcal{C})$-covers $\mathbf{K}_L$ will imply that,

$$\mathcal{I}_{PBPol(\mathcal{C})}[L, \rho] \subseteq \mathcal{I}[\rho](\mathbf{K}_L) \subseteq \{r \mid (L, r) \in S\}$$

Thus, by definition of $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$, it will follow that $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho] \subseteq S$. Before, we present this construction, we introduce some additional terminology that we shall need. More precisely, we prove a property of the set $\mathcal{T}$ which will be crucial for the construction.

9.1. **Preliminary results.** We say that a language $G \subseteq A^*$ is *good* when the two following conditions are satisfied:

- G is closed under infixes: for any $u, v_1, v_2 \in A^*$, if $v_1 u v_2 \in G$, then $u \in G$.
- $\tau(G) \subseteq S$.

Note that we already know at least one simple good language: $\emptyset$ which trivially satisfies both conditions.

Remark 9.1. *Our main goal is to prove that $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho] \subseteq S$. Thus, since $\tau(A^*) = \{(L, r) \mid r \in \mathcal{I}_{PBPol(\mathcal{C})}[L, \rho]\} = \mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho]$ by definition, this means that the universal language A^* is good. However, we* **do not know this yet***. A key point in the proof is that we exhibit increasingly large good languages until we obtain that A^* itself is good.*

We shall need the following fact which is used for combining two good languages into a single larger one.

Fact 9.2. *Let G_1, G_2 be good languages. Then $G_1G_2 \cup G_1 \cup G_2$ is a good language as well.*

Proof. Clearly, $G_1G_2 \cup G_1 \cup G_2$ is closed under infixes since this was the case for both G_1 and G_2. Hence, we may concentrate on the second property in the definition of good languages. Since τ is a multiplicative rating map, we have,

$$\tau(G_1G_2 \cup G_1 \cup G_2) = \tau(G_1) \cdot \tau(G_2) \cup \tau(G_1) \cup \tau(G_2)$$

We already know that $\tau(G_1) \subseteq S$ and $\tau(G_2) \subseteq S$ since G_1, G_2 are good. Thus, it suffices to verify that $\tau(G_1) \cdot \tau(G_2) \subseteq S$. By definition, $(L, r) \in \tau(G_1) \cdot \tau(G_2)$ satisfies $L = L_1 \odot L_2$ and $r \leq r_1 r_2$ for $(L_1, r_1) \in \tau(G_1) \subseteq S$ and $(L_2, r_2) \in \tau(G_2) \subseteq S$. Since S is $PBPol(\mathcal{C})$-saturated, it is closed under downset and multiplication which yields $(L, r) \in S$. $\square$

We use good languages to introduce a new multiplicative rating map $\gamma : 2^{A^*} \to 2^{\mathbf{L} \times R}$ which is a key ingredient of the proof. It is defined from τ, the $PBPol(\mathcal{C})$-derived multiplicative rating map associated to $\mathbf{L}$ and ρ. Recall that $2^{\mathbf{L} \times R}$ is a finite hemiring (we showed this when defining τ at the beginning of Section 8). We define,

$$\begin{array}{llll} \gamma : & 2^{A^*} & \to & 2^{\mathbf{L} \times R} \\ & K & \mapsto & \{(L, r) \mid (L, r) \in \tau(K \cap G) \text{ for some good language } G\} \end{array}$$

Remark 9.3. *This new map γ is strongly related to τ. An apparent connection is that for any $K \subseteq A^*$, we have $\gamma(K) \subseteq \tau(K)$. Indeed, given $(L, r) \in \gamma(K)$, we have $(L, r) \in \tau(K \cap G)$ for some good language G. Moreover, we know that $\tau(K \cap G) \subseteq \tau(K)$ since $K \cap G \subseteq K$. Thus, we get that $(L, r) \in \tau(K)$. However, the connection between γ and τ is in fact much stronger. Since we shall later obtain that A^* is good (see Remark 9.1), the converse inclusion holds as well: $\gamma(K) \subseteq \tau(K)$. Thus, γ and τ are actually the same object. Unfortunately, this is an information that we are not able to use. Indeed, we only obtain it at the end of our proof: the fact that A^* is good follows from our goal ($\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L}, \rho] \subseteq S$).*

Before we can use γ, we need to verify that it is indeed a multiplicative rating map. We do so in the next lemma.

Lemma 9.4. *The map γ is a multiplicative rating map.*

Proof. We have to show that γ is a hemiring morphism. We start by proving that it is a monoid morphism for addition. Clearly, $\gamma(\emptyset) = \emptyset$. Indeed,

$$\gamma(\emptyset) = \{(L, r) \mid (L, r) \in \tau(\emptyset)\} = \emptyset$$

We now show that for any $K_1, K_2 \subseteq A^*$, we have $\gamma(K_1 \cup K_2) = \gamma(K_1) \cup \gamma(K_2)$. Assume first that $(L, r) \in \gamma(K_1) \cup \gamma(K_2)$. By symmetry assume that $(L, r) \in \gamma(K_1)$. By definition, we have a good language G such that $(L, r) \in \tau(K_1 \cap G)$. Thus, since $K_1 \cap G \subseteq (K_1 \cup K_2) \cap G$, we get $(L, r) \in \tau((K_1 \cup K_2) \cap G)$ which means that $(L, r) \in \gamma(K_1 \cup K_2)$ as desired. Conversely, assume that $(L, r) \in \gamma(K_1 \cup K_2)$. By definition, there exists a good language G such that $(L, r) \in \tau((K_1 \cup K_2) \cap G)$. Thus, since τ is a morphism itself, we have $(L, r) \in \tau(K_1 \cap G) \cup \tau(K_2 \cap G)$ and it follows that $(L, r) \in \gamma(K_1) \cup \gamma(K_2)$.

It remains to prove that γ is a morphism for multiplication. Let $K_1, K_2 \subseteq A^*$. We show that $\gamma(K_1K_2) = \gamma(K_1) \cdot \gamma(K_2)$. Assume first that $(L, r) \in \gamma(K_1) \cdot \gamma(K_2)$. Thus, we have $(L_1, r_1) \in \gamma(K_1)$ and $(L_2, r_2) \in \gamma(K_2)$ such that $L = L_1 \odot L_2$ and $r \leq r_1r_2$. It then follows from the definition of γ that there exist two good languages G_1, G_2 such that,

$$(L_1, r_1) \in \tau(K_1 \cap G_1) \quad \text{and} \quad (L_2, r_2) \in \tau(K_2 \cap G_2)$$

Since τ is a morphism, it follows that $(L, r) \in \tau((K_1 \cap G_1)(K_2 \cap G_2))$. Moreover, since $(K_1 \cap G_1)(K_2 \cap G_2) \subseteq K_1K_2 \cap (G_1G_2 \cup G_1 \cup G_2)$, we get $(L, r) \in \tau(K_1K_2 \cap (G_1G_2 \cup G_1 \cup G_2))$. Finally, since we know that $G_1G_2 \cup G_1 \cup G_2$ is good by Fact 9.2, it follows that $(L, r) \in \gamma(K_1K_2)$ by definition of γ.

We finish with the converse inclusion. Let $(L, r) \in \gamma(K_1K_2)$. By definition, we get a good language G such that $(L, r) \in \tau(K_1K_2 \cap G)$. Since G is closed under infixes, it is simple to verify that $K_1K_2 \cap G \subseteq (K_1 \cap G)(K_2 \cap G)$. Thus, it follows that,

$$(L, r) \in \tau((K_1 \cap G)(K_2 \cap G)) = \tau(K_1 \cap G) \cdot \tau(K_2 \cap G)$$

Therefore, we have $(L_1, r_1) \in \tau(K_1 \cap G)$ and $(L_2, r_2) \in \tau(K_2 \cap G)$ such that $L = L_1 \odot L_2$ and $r \leq r_1r_2$. By definition, $(L_1, r_1) \in \gamma(K_1)$ and $(L_2, r_2) \in \gamma(K_2)$. Thus, we obtain that, $(L, r) \in \gamma(K_1) \cdot \gamma(K_2)$. ☐

Now that we know that γ is a multiplicative rating map, we consider $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \gamma]$, the $Pol(\mathcal{C})$-optimal $\mathbf{H}_\rho$-pointed γ-imprint. Recall that $\mathbf{H}_\rho$ is multiplicative and $\mathcal{C}$-compatible by definition. Thus, it is immediate from Theorem 7.5 that $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \gamma]$ is the smallest $Pol(\mathcal{C})$-saturated subset of $\mathbf{H}_\rho \times 2^{\mathbf{L} \times R}$ (for γ). This information yields the following proposition.

Proposition 9.5. *We have $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \gamma] \subseteq \mathcal{T}$.*

Proof. By Theorem 7.5, $\mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \gamma]$ is the smallest $Pol(\mathcal{C})$-saturated subset of $\mathbf{H}_\rho \times 2^{\mathbf{L} \times R}$ (for γ). Thus, it suffices to show that $\mathcal{T}$ is $Pol(\mathcal{C})$-saturated. Recall that by hypothesis, the pair $(S, \mathcal{T})$ is $PBPol(\mathcal{C})$-saturated. Hence, we already know that $\mathcal{T}$ is closed under downset and multiplication. It remains to show that $\mathcal{P}_{triv}[\mathbf{H}_\rho, \gamma] \subseteq \mathcal{T}$ and that $\mathcal{T}$ satisfies $Pol(\mathcal{C})$-closure.

We start with $\mathcal{P}_{triv}[\mathbf{H}_\rho, \gamma] \subseteq \mathcal{T}$. Let $(H, T) \in \mathcal{P}_{triv}[\mathbf{H}_\rho, \gamma]$, we show that $(H, T) \in \mathcal{T}$. By definition of $\mathcal{P}_{triv}[\mathbf{H}_\rho, \gamma]$, there exists $w \in H$ such that $T \subseteq \gamma(w)$. Moreover, as we observed in Remark 9.3, $\gamma(w) \subseteq \tau(w)$. Therefore, $T \subseteq \tau(w)$ and we have $(H, T) \in \mathcal{P}_{triv}[\mathbf{H}_\rho, \tau]$. This concludes the proof since we know that $\mathcal{T}$ contains $\mathcal{P}_{triv}[\mathbf{H}_\rho, \tau]$ (this is because $(S, \mathcal{T})$ is $PBPol(\mathcal{C})$-saturated).

We turn to $Pol(\mathcal{C})$-closure. Let $(E, F) \in \mathcal{T} \subseteq \mathbf{H}_\rho \times 2^{\mathbf{L} \times R}$ be an idempotent. We have to show that,

$$(E, F \cdot \gamma([E]_\mathcal{C}) \cdot F) \in \mathcal{T}$$

Let $T = \{(L, r) \in S \mid [E]_\mathcal{C} = [L]_\mathcal{C}\}$. Since $(S, \mathcal{T})$ is $PBPol(\mathcal{C})$-saturated, we get from nested closure that,

$$(E, F \cdot T \cdot F) \in \mathcal{T}$$

Hence, by closure under downset, it suffices to show that $\gamma([E]_{\mathcal{C}}) \subseteq T$. Let $(L,r) \in \gamma([E]_{\mathcal{C}})$. By definition, we know that $(L,r) \in \tau([E]_{\mathcal{C}} \cap G) \subseteq \tau(G)$ for some good language G. By the second item in the definition of good languages, this means that $(L,r) \in S$. It remains to verify that $[E]_{\mathcal{C}} = [L]_{\mathcal{C}}$. By definition of τ, $(L,r) \in \tau([E]_{\mathcal{C}} \cap G)$ means that we have $r \in \mathcal{I}_{PBPol(\mathcal{C})}[[E]_{\mathcal{C}} \cap G \cap L, \rho]$. Therefore, $[E]_{\mathcal{C}} \cap G \cap L \neq \emptyset$ and since $\mathbf{L}$ is $\mathcal{C}$-compatible, we get $[E]_{\mathcal{C}} = [L]_{\mathcal{C}}$. $\square$

9.2. **Main argument.** We now come back to the main proof. Recall that our objective is to show the inclusion $\mathcal{P}_{PBPol(\mathcal{C})}[\mathbf{L},\rho] \subseteq S$. As we explained, it suffices to present a construction which builds $PBPol(\mathcal{C})$-covers $\mathbf{K}_L$ for all $L \in \mathbf{L}$ such that for any $K \in \mathbf{K}_L$, we have $(L,\rho(K)) \in S$. This construction is based on Simon's factorization forest theorem which we presented in Section 4. Recall that using it requires a monoid morphism $\alpha : A^* \to M$. Therefore, we start by defining α.

For any language $H \in \mathbf{H}_\rho$, we let $\mathbf{P}_H$ as an arbitrary *optimal* $Pol(\mathcal{C})$-cover of H for the multiplicative rating map γ defined above. Let $\mathbf{P} = \cup_{H \in \mathbf{H}_\rho} \mathbf{P}_H$. By definition, $\mathbf{P}$ is a finite set of languages in $Pol(\mathcal{C})$. Therefore, there exists a stratum $Pol_k(\mathcal{C})$ in our stratification of $Pol(\mathcal{C})$ (defined in Section 4, recall that $\mathcal{C}$ is finite) such that all $P \in \mathbf{P}$ belong to $Pol_k(\mathcal{C})$. We define $\mathcal{D} = Bool(Pol_k(\mathcal{C}))$. By definition, $\mathcal{D}$ is a finite quotienting Boolean algebra and any $P \in \mathbf{P}$ belongs to $\mathcal{D}$. Consider, the canonical equivalence $\sim_{\mathcal{D}}$ (over A^*) associated to $\mathcal{D}$. Since $\mathcal{D}$ is quotienting, it follows from Lemma 2.8 that $\sim_{\mathcal{D}}$ is a congruence of finite index for word concatenation. Therefore the quotient set $A^*/{\sim_{\mathcal{D}}}$ is a finite monoid and the map $w \mapsto [w]_{\mathcal{D}}$ (which associates its $\sim_{\mathcal{D}}$-class to any word) is a morphism.

Moreover, recall that since $\mathbf{L}$ is multiplicative, the powerset $2^{\mathbf{L}}$ is a semigroup: given $\mathbf{L}_1, \mathbf{L}_2 \in 2^{\mathbf{L}}$, their multiplication is,

$$\mathbf{L}_1 \odot \mathbf{L}_2 = \{L_1 \odot L_2 \mid L_1 \in \mathbf{L}_1 \text{ and } L_2 \in \mathbf{L}_2\}$$

Therefore, the Cartesian product $2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})$ is a semigroup for the componentwise multiplication. We define a morphism $\alpha : A^* \to 2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})$ as follows. For any $w \in A^*$, we let,

$$\alpha(w) = (\{L \in \mathbf{L} \mid w \in L\}, [w]_{\mathcal{D}})$$

One may verify that α is indeed a (semigroup) morphism. Thus, the restriction of α to its image is a monoid morphism as desired. We work with α-factorization forests. Observe that any element $s \in 2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})$ is a pair. We write it $s = (\mathbf{L}_s, V_s)$ (i.e. $\mathbf{L}_s \in 2^{\mathbf{L}}$ and $V_s \in A^*/{\sim_{\mathcal{D}}}$).

We are now ready to start our construction. We state it in the following proposition. Recall that for any $s \in 2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})$ and any $h,p \in \mathbb{N}$, $F^\alpha(s,h,p) \subseteq A^*$ denotes the language of all words in $\alpha^{-1}(s)$ admitting an α-factorization forest of height at most h and idempotent height at most p.

Proposition 9.6. *Let $h,p \in \mathbb{N}$ and $s \in 2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})$. There exists a $PBPol(\mathcal{C})$-cover $\mathbf{K}$ of $F^\alpha(s,h,p)$ such that for any $K \in \mathbf{K}$ and $L \in \mathbf{L}_s$, we have $(L,\rho(K)) \in S$.*

Before proving the proposition, we use it to finish the completeness proof. Consider a language $L \in \mathbf{L}$. Our objective is to build a $PBPol(\mathcal{C})$-cover $\mathbf{K}_L$ of L such that,

$$(L,\rho(K)) \in S \quad \text{for all } K \in \mathbf{K}_L$$

Let $h = p = 3|2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})| - 1$. We obtain from Simon's factorization forest theorem (i.e. Theorem 4.1) that for any $s \in 2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})$, we have $\alpha^{-1}(s) = F^{\alpha}(s, h, p)$. Thus, given any $s \in 2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})$, Proposition 9.6 yields a $PBPol(\mathcal{C})$-cover $\mathbf{K}_s$ of $\alpha^{-1}(s)$ such that for any $K \in \mathbf{K}_s$ and $L \in \mathbf{L}_s$, we have $(L, \rho(K)) \in S$. Observe that by definition of α, we have,

$$L = \bigcup_{\{s \in 2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}}) \mid L \in \mathbf{L}_s\}} \alpha^{-1}(s)$$

Therefore, it suffices to define $\mathbf{K}_L$ as the union of all sets $\mathbf{K}_s$ such that $L \in \mathbf{L}_s$. It is immediate by definition that $\mathbf{K}_L$ is a $PBPol(\mathcal{C})$-cover of L and that $(L, \rho(K)) \in S$ for all $K \in \mathbf{K}_L$.

It remains to prove Proposition 9.6. Let $h, p \in \mathbb{N}$ and $s \in 2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})$. Our objective is to build a $PBPol(\mathcal{C})$-cover $\mathbf{K}$ of $F^{\alpha}(s, h, p)$ which satisfies the following property:

$$\text{For all } K \in \mathbf{K} \text{ and any } L \in \mathbf{L}_s, \quad (L, \rho(K)) \in S \tag{9.1}$$

Observe that if $\alpha^{-1}(s) = \emptyset$, this is easily achieved: it suffices to define $\mathbf{K} = \emptyset$. Thus, we shall assume from now one that $\alpha^{-1}(s) \neq \emptyset$. The construction itself is very similar to the one we used when proving Proposition 7.8 (i.e. the $Pol(\mathcal{C})$ construction). However, showing that this second construction satisfies (9.1) will require a lot more work in this case. Moreover, we proceed by induction on *two* parameters which are listed below by order of importance:

(1) The idempotent height p.
(2) The height h.

We start with the induction base: $h = 0$. Let us point out that for this case, the argument is actually slightly simpler than the one we used for the corresponding case in the $Pol(\mathcal{C})$ proof. This is because $PBPol(\mathcal{C})$ contains all finite languages (this may or may not be the case for $Pol(\mathcal{C})$ depending on the finite quotienting Boolean algebra $\mathcal{C}$).

Base case: Leaves. Assume that $h = 0$. All words in $F^{\alpha}(s, 0, p)$ are either empty or made of a single letter $a \in A$. In particular $F^{\alpha}(s, 0, p)$ is a *finite language*. Thus, the following set $\mathbf{K}$ is finite:

$$\mathbf{K} = \{\{w\} \mid w \in F^{\alpha}(s, 0, p)\}$$

Clearly, $\mathbf{K}$ is a finite cover of $F^{\alpha}(s, 0, p)$. Moreover, it is a $PBPol(\mathcal{C})$-cover since we know that $PBPol(\mathcal{C})$ contains all finite languages (see Lemma 3.4). It remains to show that (9.1) is satisfied.

Let $K \in \mathbf{K}$ and $L \in \mathbf{L}_s$, we show that $(L, \rho(K)) \in S$. Since $K \in \mathbf{K}$, we know that $K = \{w\}$ for some $w \in F^{\alpha}(s, 0, p)$. In particular, we have $\alpha(w) = s$ which means that $w \in L$ since $L \in \mathbf{L}_s$. Thus, we have,

$$(L, \rho(K)) = (L, \rho(w)) \in \mathcal{P}_{triv}[\mathbf{L}, \rho]$$

This concludes the proof. Indeed, we know that S is $PBPol(\mathcal{C})$-saturated which means that $\mathcal{P}_{triv}[\mathbf{L}, \rho] \subseteq S$ by definition.

Inductive case. We now assume that $h \geq 1$. Recall that our objective is to build a $PBPol(\mathcal{C})$-cover $\mathbf{K}$ of $F^{\alpha}(s, h, p)$ which satisfies (7.1). The construction is similar to the one used for $Pol(\mathcal{C})$ when proving Proposition 7.8 in Section 7: we decompose $F^{\alpha}(s, h, p)$ as the union of three languages that we cover independently.

Recall that $F^{\alpha}_B(s, h, p)$ (resp. $F^{\alpha}_I(s, h, p)$) denotes the language of all words in $\alpha^{-1}(s)$ admitting an α-factorization forest of height of at most h, of idempotent height at most p

and whose root is a *binary node* (resp. *idempotent node*). The construction is based on the two following lemmas.

Lemma 9.7. *There exists a $PBPol(\mathcal{C})$-cover $\mathbf{K}_B$ of $F^\alpha_B(s,h,p)$ such that for all $K \in \mathbf{K}_B$ and any $L \in \mathbf{L}_s$, we have $(L,\rho(K)) \in S$.*

Lemma 9.8. *There exists a $PBPol(\mathcal{C})$-cover $\mathbf{K}_I$ of $F^\alpha_I(s,h,p)$ such that for all $K \in \mathbf{K}_I$ and any $L \in \mathbf{L}_s$, we have $(L,\rho(K)) \in S$.*

Before we show these two results, let us use them to finish the proof. Let $\mathbf{K}_B$ and $\mathbf{K}_I$ be as defined in the lemma. Observe that since we assumed that $h \geq 1$ it is immediate that $F^\alpha(s,h,p)$ is equal to the following union:

$$F^\alpha(s,h,p) = F^\alpha_B(s,h,p) \cup F^\alpha_I(s,h,p) \cup F^\alpha(s,h-1,p)$$

Moreover, induction on h yields a $PBPol(\mathcal{C})$-cover $\mathbf{K}'$ of $F^\alpha(s,h-1,p)$ such that for all $K \in \mathbf{K}$ and any $L \in \mathbf{L}_s$, we have $(L,\rho(K)) \in S$. Thus, it suffices to define $\mathbf{K} = \mathbf{K}_B \cup \mathbf{K}_I \cup \mathbf{K}'$. By definition, we know that $\mathbf{K}$ is a $PBPol(\mathcal{C})$-cover of $F^\alpha(s,h,p)$ which satisfies (7.1).

It remains to prove the two lemmas. We start with Lemma 9.7 which is simpler. The argument is essentially identical to the one we used for proving the corresponding result for $Pol(\mathcal{C})$ in Section 7.

9.3. **Proof of Lemma 9.7.** For all $t \in 2^{\mathbf{L}} \times (A^*/{\sim_\mathcal{D}})$, consider the language $F^\alpha(t,h-1,p)$. Using induction on our second parameter (the height h), we get that for all $t \in 2^{\mathbf{L}} \times (A^*/{\sim_\mathcal{D}})$, there exists a $PBPol(\mathcal{C})$-cover $\mathbf{U}_t$ of $F^\alpha(t,h-1,p)$ such that for any $U \in \mathbf{U}_t$ and any $L \in \mathbf{L}_t$, we have $(L,\rho(U)) \in S$. Consider the following set:

$$\mathbf{K}_B = \{K_1K_2 \mid \text{there exists } t_1,t_2 \in 2^{\mathbf{L}} \times (A^*/{\sim_\mathcal{D}}) \text{ s.t. } s = t_1t_2,\ K_1 \in \mathbf{U}_{t_1} \text{ and } K_2 \in \mathbf{U}_{t_2}\}$$

One may now verify that $\mathbf{K}_B$ is a $PBPol(\mathcal{C})$-cover of $F^\alpha(s,h,p)$ and that we have $(L,\rho(K)) \in S$ for $K \in \mathbf{K}_B$ and $L \in \mathbf{L}_s$. The proof is left to the reader as it is identical to that of Lemma 7.11 in Section 7. In particular, when proving that $(L,\rho(K)) \in S$ for $K \in \mathbf{K}_B$ and $L \in \mathbf{L}_s$, one uses the fact that S is $PBPol(\mathcal{C})$-saturated (specifically closure under multiplication).

9.4. **Proof of Lemma 9.8.** We turn to Lemma 9.8. Our objective is to construct a $PBPol(\mathcal{C})$-cover $\mathbf{K}_I$ of $F^\alpha_I(s,h,p)$ such that for all $K \in \mathbf{K}_I$ and $L \in \mathbf{L}_s$, we have $(L,\rho(K)) \in S$. Note that we may assume without loss of generality that $p \geq 1$ and s is an idempotent of $2^{\mathbf{L}} \times (A^*/{\sim_\mathcal{D}})$. Indeed, otherwise we have $F^\alpha_I(s,h,p) = \emptyset$ and we may simply choose $\mathbf{K}_I = \emptyset$. We shall write $s = e \in 2^{\mathbf{L}} \times (A^*/{\sim_\mathcal{D}})$ to underline the fact that s is idempotent: we have to cover $F^\alpha_I(e,h,p)$.

This proof is where the argument departs from what we did for $Pol(\mathcal{C})$ in Section 7. On one hand, the construction is still based on Lemma 7.2. However, there is a subtle difference and showing that $(L,\rho(K)) \in S$ for all $K \in \mathbf{K}_I$ and $L \in \mathbf{L}_e$ is much more involved (this is where we shall use induction on the idempotent height p).

Recall that applying Lemma 7.2 requires a cover $\mathbf{U}$ of $F^\alpha(e,h-1,p-1)$ and a language V containing $\alpha^{-1}(e)$. We first define these two objects. We define $\mathbf{U}$ as an optimal $PBPol(\mathcal{C})$-cover of $F^\alpha_I(e,h-1,p-1)$. Moreover, recall that $e \in 2^{\mathbf{L}} \times (A^*/{\sim_\mathcal{D}})$ is a pair $e = (\mathbf{L}_e, V_e)$ where V_e is a $\sim_\mathcal{D}$-class. Clearly, we have $\alpha^{-1}(e) \subseteq V_e$ (indeed $V_e = [w]_\mathcal{D}$ for all $w \in \alpha^{-1}(e)$ by definition of α).

Remark 9.9. *There is a subtle but crucial difference with what we did for $Pol(\mathcal{C})$. The $PBPol(\mathcal{C})$-cover $\mathbf{U}$ is* **not** *obtained from induction. Instead, it is an arbitrary optimal $PBPol(\mathcal{C})$-cover of $F_I^\alpha(e,h-1,p-1)$. We shall use induction later for a different purpose.*

Now that we have $\mathbf{U}$ and V_e, we may apply Lemma 7.2. We obtain a cover $\mathbf{K}_I$ of $F_I^\alpha(e,h,p)$ such that any $K \in \mathbf{K}_I$ is a concatenation $K = K_1 \cdots K_n$ where each K_i is of one of the two following kinds:

(1) K_i is a language in $\mathbf{U}$, or,
(2) $K_i = U_1 \cdots U_m V_e U'_1 \cdots U'_{m'}$ where $U_1, \cdots, U_m, U'_1, \dots, U'_{m'} \in \mathbf{U}$ and there exists an *idempotent* $f \in R$ such that $\rho(U_1 \cdots U_m) = \rho(U'_1 \cdots U'_{m'}) = f$.

Clearly, any $K \in \mathbf{K}_I$ belongs to $PBPol(\mathcal{C})$. Indeed, K is a concatenation of languages in $PBPol(\mathcal{C})$ by definition ($V_e \in PBPol(\mathcal{C})$ as it is $\sim_\mathcal{D}$-class and by construction, we have $k \in \mathbb{N}$ such that $\mathcal{D} = Bool(Pol_k(\mathcal{C})) \subseteq PBPol(\mathcal{C})$). Therefore, $\mathbf{K}_I$ is a $PBPol(\mathcal{C})$-cover of $F_I^\alpha(e,h,p)$.

It remains to prove that for any $K \in \mathbf{K}_I$ and $L \in \mathbf{L}_e$, we have $(L,\rho(K)) \in S$. We use the two following lemmas (which are obtained from induction on our two parameters).

Lemma 9.10. *For any $U \in \mathbf{U}$ and $L \in \mathbf{L}_e$, we have $(L,\rho(U)) \in S$.*

Lemma 9.11. *Let $f \in R$ be an idempotent such that $f = \rho(U_1 \cdots U_m)$ with $U_1, \dots, U_m \in \mathbf{U}$. For any $L \in \mathbf{L}_e$, we have $(L, f \cdot \rho(V_e) \cdot f) \in S$.*

Let us first apply these two lemmas and finish the proof of Lemma 9.8. Let $K \in \mathbf{K}_I$ and $L \in \mathbf{L}_e$, we show that $(L,\rho(K)) \in S$. By definition of $\mathbf{K}_I$, we know that $K = K_1 \cdots K_n$ where all languages K_i are as described in the two items above. Moreover, since $e = (\mathbf{L}_e, V_e)$ is idempotent, we know that $\mathbf{L}_e \in 2^\mathbf{L}$ is idempotent as well. Hence, $\mathbf{L}_e = (\mathbf{L}_e)^n$ and since $L \in \mathbf{L}_e$, we get $L_1, \dots, L_n \in \mathbf{L}_e$ such that $L = L_1 \odot \cdots \odot L_n$.

Observe that for all $i \leq n$, we have $(L_i, \rho(K_i)) \in S$. If K_i is as described in the first item ($K_i \in \mathbf{U}$), this is by Lemma 9.10. Otherwise, when K_i is as described in the second item, this is by Lemma 9.11. Therefore, since S is $Pol(\mathcal{C})$-saturated, we obtain from closure under multiplication that,

$$(L,\rho(K)) = (L_1 \odot \cdots \odot L_n, \rho(K_1) \cdots \rho(K_n)) \in S$$

This concludes the proof of Lemma 9.8. It remains to show Lemmas 9.10 and Lemma 9.11. We start with the former which is simpler.

9.4.1. *Proof of Lemma 9.10.* Consider $U \in \mathbf{U}$ and $L \in \mathbf{L}_e$. We show that $(L,\rho(U)) \in S$. Recall that $\mathbf{U}$ is an optimal $PBPol(\mathcal{C})$-cover of $F_I^\alpha(e,h-1,p-1)$ by definition. Thus, we have,

$$\rho(U) \in \mathcal{I}[\rho](\mathbf{U}) = \mathcal{I}_{PBPol(\mathcal{C})}[F_I^\alpha(e,h-1,p-1),\rho]$$

Moreover, using induction (we may use any of our two parameters here since they have both decreased), we get a $PBPol(\mathcal{C})$-cover $\mathbf{U}'$ of $F_I^\alpha(e,h-1,p-1)$ such that $(L',\rho(U')) \in S$ for any $L' \in \mathbf{L}_e$ (such as L) and any $U' \in \mathbf{U}'$. Since $\mathbf{U}'$ is a $PBPol(\mathcal{C})$-cover of $F_I^\alpha(e,h-1,p-1)$, we know by definition that,

$$\mathcal{I}_{PBPol(\mathcal{C})}[F_I^\alpha(e,h-1,p-1),\rho] \subseteq \mathcal{I}[\rho](\mathbf{U}')$$

Altogether, this means that $\rho(U) \in \mathcal{I}[\rho](\mathbf{U}')$ and by definition, this yields $U' \in \mathbf{U}'$ such that $\rho(U) \leq \rho(U')$. In particular, we have $(L,\rho(U')) \in S$ by definition of $\mathbf{U}'$. Thus, since S is $PBPol(\mathcal{C})$-saturated, closure under downset yields $(L,\rho(U)) \in S$.

9.4.2. *Proof of Lemma 9.11.* Consider an idempotent $f \in R$ such that $f = \rho(U_1 \cdots U_m)$ for some $U_1, \dots, U_m \in \mathbf{U}$. We show that for any $L \in \mathbf{L}_e$, we have $(L, f \cdot \rho(V_e) \cdot f) \in S$. Let us start with a simpler result which we prove with Lemma 9.10.

Fact 9.12. *For any $L \in \mathbf{L}_e$, we have $(L, f) \in S$.*

Proof. Since e is idempotent, this is also the case for $\mathbf{L}_e \in 2^{\mathbf{L}}$. Thus, $\mathbf{L}_e = (\mathbf{L}_e)^m$ and we have $L_1, \dots, L_m \in \mathbf{L}_e$ such that $L = L_1 \odot \cdots \odot L_m$. Using Lemma 9.10, we obtain that $(L_i, \rho(U_i)) \in S$ for all $i \leq m$. Thus, since S is $PBPol(\mathcal{C})$-saturated, we obtain from closure under multiplication that $(L, f) = (L_1 \odot \cdots \odot L_m, \rho(U_1 \cdots U_m)) \in S$. □

We now come back to the main proof. Let $L \in \mathbf{L}_e$. We show that $(L, f \cdot \rho(V_e) \cdot f) \in S$. We first prove that we may restrict ourselves to the special case when L is an idempotent without loss of generality.

Recall that since e is an idempotent of $2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})$, $\mathbf{L}_e$ is an idempotent of $2^{\mathbf{L}}$. Thus, we have $\mathbf{L}_e = (\mathbf{L}_e)^{|\mathbf{L}|+1}$, and we get $L_1, \dots, L_{|\mathbf{L}|+1} \in \mathbf{L}_e$ such that, $L = L_1 \odot \cdots \odot L_{|\mathbf{L}|+1}$. One may then use a pumping argument to obtain $L', L'', E \in \mathbf{L}_e$ such that E is idempotent and $L = L' \odot E \odot L''$. Using Fact 9.12, we get that $(L', f), (L'', f) \in S$. We now prove that the following property also holds,

$$(E, f \cdot \rho(V_e) \cdot f) \in S \tag{9.2}$$

Since S is $PBPol(\mathcal{C})$-saturated, it will then follow by closure under multiplication that,

$$(L, f \cdot \rho(V_e) \cdot f) = (L' \odot E \odot L'', f \cdot f \cdot \rho(V_e) \cdot f \cdot f) \in S$$

It remains to prove that (9.2) holds. We use $PBPol(\mathcal{C})$-closure and the next lemma (which is where we apply induction on the idempotent height p).

Lemma 9.13. *Let $v \in V_e$ and $H = \{w \in A^* \mid \rho(w) = \rho(v)\} \in \mathbf{H}_\rho$. There exists $T \in 2^{\mathbf{L} \times R}$ such that $(H, T) \in \mathcal{T}$ and $(E, f) \in T$.*

Let us first use Lemma 9.13 to finish proving that (9.2) holds. Since ρ is **nice**, we know that there exist words $v_1, \dots, v_\ell \in V_e$ such that, $\rho(v_1) + \cdots + \rho(v_\ell) = \rho(V_e)$. Therefore, we have the following fact.

Fact 9.14. *The following inequality holds,*

$$f \cdot \rho(V_e) \cdot f \leq \prod_{1 \leq i \leq \ell} (f \cdot (\rho(v_i) + \rho(\varepsilon)) \cdot f)$$

Proof. Since $f = \rho(U_1 \cdots U_m)$, we have $\rho(\varepsilon) \cdot f = f$. Thus, we may distribute multiplications to obtain that,

$$\sum_{1 \leq i \leq \ell} f \cdot \rho(v_i) \cdot f \leq \prod_{1 \leq i \leq \ell} (f \cdot (\rho(v_i) + \rho(\varepsilon)) \cdot f)$$

Finally, since $\rho(v_1) + \cdots + \rho(v_\ell) = \rho(V_e)$, it is immediate that we have,

$$f \cdot \rho(V_e) \cdot f = \sum_{1 \leq i \leq \ell} f \cdot \rho(v_i) \cdot f$$

This concludes the proof of Fact 9.14. □

For all $i \leq \ell$, let $H_i = \{w \in A^* \mid \rho(w) = \rho(v_i)\} \in \mathbf{H}_\rho$. It follows from Lemma 9.13 that for all $i \leq \ell$, there exists $T_i \in 2^{\mathbf{L} \times R}$ such that $(H_i, T_i) \in \mathcal{T}$ and $(E, f) \in T_i$. Therefore, we may use the fact that $(S, \mathcal{T})$ is $PBPol(\mathcal{C})$-saturated (specifically $PBPol(\mathcal{C})$-closure) to obtain that for any $i \leq \ell$, we have,

$$(E, f \cdot (\rho(v_i) + \rho(\varepsilon)) \cdot f) \in S$$

Since E is idempotent, we may now use closure under multiplication to obtain,

$$\left(E, \prod_{1 \leq i \leq \ell} f \cdot (\rho(v_i) + \rho(\varepsilon)) \cdot f\right) \in S$$

Finally, using Fact 9.14 and closure under downset, we obtain as desired that we have,

$$(E, f \cdot \rho(V_e) \cdot f) \in S$$

Hence, we get that (9.2) holds which concludes the proof of Lemma 9.11.

We turn to the proof of Lemma 9.13. The remainder of the section is devoted to it. Recall that $E \in \mathbf{L}_e$ is an idempotent and consider $v \in V_e$. Let $H = \{w \in A^* \mid \rho(w) = \rho(v)\} \in \mathbf{H}_\rho$. We show that there exists $T \in 2^{\mathbf{L} \times R}$ such that $(H, T) \in \mathcal{T}$ and $(E, f) \in T$.

Recall that we defined an optimal $Pol(\mathcal{C})$-cover $\mathbf{P}_H$ of the language $H \in \mathbf{H}_\rho$ for the multiplicative rating map γ defined at the beginning of the section (we used $\mathbf{P}_H$ to build the finite quotienting Boolean algebra $\mathcal{D}$: it is chosen so that any $P \in \mathbf{P}_H$ belongs to $\mathcal{D}$). Thus, since $v \in H$, there exists $P \in \mathbf{P}_H$ such that $v \in P$. Observe that since $v \in V_e$, we have the following fact:

Fact 9.15. *We have $V_e \subseteq P$.*

Proof. We know that $P \in \mathcal{D}$ and V_e is a $\sim_\mathcal{D}$-class. Thus, since v belongs to both V_e and P, it follows that $V_e \subseteq P$. □

We choose $T = \gamma(P) \in 2^{\mathbf{L} \times R}$. It remains to show that $(H, T) \in \mathcal{T}$ and $(E, f) \in T$. We start with $(H, T) \in \mathcal{T}$. Since $\mathbf{P}_H$ is an optimal $Pol(\mathcal{C})$-cover of $H \in \mathbf{H}_\rho$ for γ, we know that,

$$(H, \gamma(P)) \in \mathcal{P}_{Pol(\mathcal{C})}[\mathbf{H}_\rho, \tau]$$

Thus, it is immediate from Proposition 9.5 that we have $(H, \gamma(P)) \in \mathcal{T}$. We now prove that $(E, f) \in T$. Since $T = \gamma(P)$, the definition of γ yields that,

$$T = \{(L, r) \in \mathbf{L} \times R \mid (L, r) \in \tau(P \cap G) \text{ for some good language } G\}$$

Therefore, we have to exhibit a good language G such that $(E, f) \in \tau(P \cap G)$. Recall that $f = \rho(U_1 \cdots U_m)$ where $U_1, \dots, U_m \in \mathbf{U}$. We define G as the language of all words which are an infix of some other word in $F^\alpha(e, g-1, h+m-2)$. We use induction on our first parameter (the idempotent height p) to prove that G is good.

Fact 9.16. *The language G is good.*

Proof. We know by definition that G is closed under infixes. Thus, we may concentrate on the second item in the definition of good languages. We show that $\tau(G) \subseteq S$. By definition, all words in G are an infix of some other word admitting an α-factorization forest of height at most $h + m - 2$ and of idempotent height at most $p - 1$. Therefore, it follows from

Proposition 4.2 that all words in G admit an α-factorization forest of height at most $h+m$ and idempotent height at most $p-1$. Thus, we have,

$$G \subseteq \bigcup_{t \in 2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})} F^\alpha(t, h+m, p-1)$$

Since τ is a multiplicative rating map, it follows that,

$$\tau(G) \subseteq \bigcup_{t \in 2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})} \tau(F^\alpha(t, h+m, p-1))$$

Therefore, it suffices to show that for all $t \in 2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})$, we have $\tau(F^\alpha(t, h+m, p-1)) \subseteq S$. Let $t \in 2^{\mathbf{L}} \times (A^*/{\sim_{\mathcal{D}}})$ and consider $(L, r) \in \tau(F^\alpha(t, h+m, p-1))$. We show that $(L, r) \in S$.

By definition of τ, we have $r \in \mathcal{I}_{PBPol(\mathcal{C})}[F^\alpha(t, h+m, p-1) \cap L, \rho]$. In particular, it follows that $F^\alpha(t, h+m, p-1) \cap L \neq \emptyset$. Thus, $\alpha^{-1}(t) \cap L \neq \emptyset$ and $L \in \mathbf{L}_t$. In particular, this means that $F^\alpha(t, h+m, p-1) \subseteq \alpha^{-1}(t) \subseteq L$ by definition of α. Altogether, we obtain,

$$r \in \mathcal{I}_{PBPol(\mathcal{C})}[F^\alpha(t, h+m, p-1), \rho]$$

Using induction on the idempotent height p (our most important parameter), we get a $PBPol(\mathcal{C})$-cover $\mathbf{K'}$ of $F^\alpha(t, h+m, p-1)$ such that for any $K' \in \mathbf{K'}$, and any $L' \in \mathbf{L}_t$ (such as L), we have $(L', \rho(K')) \in S$. Since $\mathbf{K'}$ is a $PBPol(\mathcal{C})$-cover of $F^\alpha(t, h+m, p-1)$,

$$\mathcal{I}_{PBPol(\mathcal{C})}[F^\alpha(t, h+m, p-1), \rho] \subseteq \mathcal{I}[\rho](\mathbf{K'})$$

Therefore, we have $r \in \mathcal{I}[\rho](\mathbf{K'})$ and we obtain $K' \in \mathbf{K'}$ such that $r \leq \rho(K')$. Finally, since $L \in \mathbf{L}_t$, we have $(L, \rho(K')) \in S$ by definition of $\mathbf{K'}$. Since S is $PBPol(\mathcal{C})$-saturated, closure under downset then yields $(L, r) \in S$. $\square$

It remains to show that $(E, f) \in \tau(P \cap G)$. Recall that $\mathbf{U}$ is an optimal $PBPol(\mathcal{C})$-cover of $F^\alpha(e, h-1, p-1)$ (for ρ). Therefore, we know that for all $i \leq m$, we have,

$$\rho(U_i) \in \mathcal{I}_{PBPol(\mathcal{C})}[F^\alpha(e, h-1, p-1), \rho]$$

Moreover, since $E \in \mathbf{L}_e$, $F^\alpha(e, h-1, p-1) \subseteq E$ and $F^\alpha(e, h-1, p-1) \cap E = F^\alpha(e, h-1, p-1)$. By definition of τ, it follows that for all $i \leq m$, we have $(E, \rho(U_i)) \in \tau(F^\alpha(e, h-1, p-1))$. Therefore, we obtain that,

$$(E, f) = (E, \rho(U_1 \cdots U_m)) \in (\tau(F^\alpha(e, h-1, p-1)))^m = \tau((F^\alpha(e, h-1, p-1))^m)$$

We prove that $(F^\alpha(e, h-1, p-1))^m \subseteq P \cap G$ since τ is a multiplicative rating map, it will then follow that $(E, f) \in \tau(P \cap G)$, finishing the proof.

Consider a word $w \in (F^\alpha(e, h-1, p-1))^m$. We prove that $w \in P \cap G$. By definition, $w = w_1 \cdots w_m$ with $w_i \in F^\alpha(e, h-1, p-1)$ for all $i \leq m$. Thus, since e is idempotent, we have $\alpha(w) = e$ and w admits an α-factorization forest of height at most $h-1+m-1 = h+m-2$ and idempotent height at most $p-1$ (one may combine the m forests of $w_1, \dots, w_m$ into a single one for w using $m-1$ binary nodes). Altogether, we obtain,

$$w \in F^\alpha(e, g-1, h+m-2)$$

Thus, it is immediate that $w \in G$ by definition of G. Moreover, we have $w \in \alpha^{-1}(e) \subseteq V_e$ and we know that $V_e \subseteq P$ by Fact 9.15. Thus, $w \in P$ which concludes the proof.

10. Conclusion

We showed that the separation and covering problems are decidable for all classes of the form $Pol(\mathcal{C})$ and $PBPol(\mathcal{C})$ where $\mathcal{C}$ is an arbitrary finite quotienting Boolean algebra. As we explained, these results have important consequences for the quantifier alternation hierarchies within first-order logic overs words: we get that separation and covering are decidable for the level $\Sigma_3(<)$. This result may be lifted to the stronger logic $\Sigma_3(<, +1, min, max, \varepsilon)$ with a transfer result of [PZ15a]. Finally, one also gets that the membership problem is decidable for $\Sigma_4(<)$ and $\Sigma_4(<, +1, min, max, \varepsilon)$ using a reduction theorem of [PZ14].

Let us point out that while this is not apparent in the above summary, we still know more about $Pol(\mathcal{C})$ than we do about $PBPol(\mathcal{C})$. Indeed, our $Pol(\mathcal{C})$-covering algorithm is based on a characterization of $Pol(\mathcal{C})$-*optimal pointed imprints* which holds for **any** multiplicative rating map. This result is stronger than the decidability of $Pol(\mathcal{C})$-covering (one only needs to consider nice multiplicative rating maps for this) and the fact that we have it is precisely why we are able to get results for the larger class $PBPol(\mathcal{C})$. On the other hand, while we also have a characterization of $PBPol(\mathcal{C})$-optimal pointed imprints, it only holds for nice multiplicative rating maps. Therefore, a natural next move is trying to generalize this characterization to all *all* multiplicative rating maps. Indeed, such a result could be the key to solving covering for one more level in concatenation hierarchies.

Another interesting (and much simpler) objective is analyzing the complexity of the decision problems that we have just solved. Naturally, this depends on the finite quotienting Boolean algebra $\mathcal{C}$ for both $Pol(\mathcal{C})$ and $PBPol(\mathcal{C})$. We leave this for further work.

References

[Arf87] 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, 1987.

[BC71] Janusz A. Brzozowski and Rina S. Cohen. Dot-depth of star-free events. *Journal of Computer and System Sciences*, 5(1):1–16, 1971.

[BK78] Janusz A. Brzozowski and Robert Knast. The dot-depth hierarchy of star-free languages is infinite. *Journal of Computer and System Sciences*, 16(1):37–55, 1978.

[Boj09] Mikołaj Bojańczyk. Factorization forests. In *Proceedings of the 13th International Conference on Developments in Language Theory*, DLT'09, pages 1–17, 2009.

[CMM13] Wojciech Czerwiński, Wim Martens, and Tomáš Masopust. Efficient separability of regular languages by subsequences and suffixes. In *Proceedings of the 40th International Colloquium on Automata, Languages, and Programming*, ICALP'13, pages 150–161, 2013.

[Col10] Thomas Colcombet. Factorization forests for infinite words and applications to countable scattered linear orderings. *Theoritical Computer Science*, 411(4-5):751–764, 2010.

[DGK08] Volker Diekert, Paul Gastin, and Manfred Kufleitner. A survey on small fragments of first-order logic over finite words. *International Journal on Foundations of Computer Science*, 19(3), 2008.

[GS00] Christian Glaßer and Heinz Schmitz. Languages of dot-depth 3/2. In *Proceedings of the 17th Annual Symposium on Theoretical Aspects of Computer Science*, (STACS'00), pages 555–566, 2000.

[Kna83] Robert Knast. A semigroup characterization of dot-depth one languages. *RAIRO - Theoretical Informatics and Applications*, 17(4):321–330, 1983.

[Kuf08] Manfred Kufleitner. The height of factorization forests. In *Proceedings of the 33rd International Symposium on Mathematical Foundations of Computer Science*, MFCS'08, pages 443–454, 2008.

[MP71] Robert McNaughton and Seymour A. Papert. *Counter-Free Automata*. MIT Press, 1971.

[Pin98] Jean-Éric Pin. Bridges for concatenation hierarchies. In *Proceedings of the 25th International Colloquium on Automata, Languages and Programming*, ICALP'98, pages 431–442, 1998.

[Pin17] Jean-Éric Pin. The dot-depth hierarchy, 45 years later. In *The Role of Theory in Computer Science. Essays Dedicated to Janusz Brzozowski*. World Scientific, 2017.

[Pla15] 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.

[PP86] Dominique Perrin and Jean-Éric Pin. First-order logic and star-free sets. *Journal of Computer and System Sciences*, 32(3):393–406, 1986.

[PS85] Jean-Éric Pin and Howard Straubing. Monoids of upper triangular boolean matrices. In *Semigroups. Structure and Universal Algebraic Problems*, volume 39, pages 259–272. North-Holland, 1985.

[PvRZ13] 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, 2013.

[PW97] Jean-Éric Pin and Pascal Weil. Polynomial closure and unambiguous product. *Theory of Computing Systems*, 30(4):383–422, 1997.

[PZ14] 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, 2014.

[PZ15a] Thomas Place and Marc Zeitoun. Separation and the successor relation. In *32nd International Symposium on Theoretical Aspects of Computer Science*, STACS'15, pages 662–675, 2015.

[PZ15b] Thomas Place and Marc Zeitoun. The tale of the quantifier alternation hierarchy of first-order logic over words. *SIGLOG news*, 2(3):4–17, 2015.

[PZ16] 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, 2016.

[PZ17a] Thomas Place and Marc Zeitoun. Concatenation hierarchies: New bottle, old wine. In *12th International Computer Science Symposium in Russia*, CSR'17, pages 25–37, 2017.

[PZ17b] Thomas Place and Marc Zeitoun. The covering problem. Unpublished, 2017.

[PZ17c] Thomas Place and Marc Zeitoun. Separation for dot-depth two. In *32th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS'17, 2017.

[Sch65] Marcel Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8(2):190–194, 1965.

[Sim75] Imre Simon. Piecewise testable events. In *Proceedings of the 2nd GI Conference on Automata Theory and Formal Languages*, pages 214–222, 1975.

[Sim90] Imre Simon. Factorization forests of finite height. *Theoritical Computer Science*, 72(1):65–94, 1990.

[Str81] Howard Straubing. A generalization of the schützenberger product of finite monoids. *Theoretical Computer Science*, 13(2):137–150, 1981.

[Str85] Howard Straubing. Finite semigroup varieties of the form V $*$ D. *Journal of Pure and Applied Algebra*, 1985.

[Str88] Howard Straubing. Semigroups and languages of dot-depth two. *Theoretical Computer Science*, 58(13):361–378, 1988.

[Thé81] Denis Thérien. Classification of finite monoids: The language approach. *Theoretical Computer Science*, 14(2):195–208, 1981.

[Tho82] Wolfgang Thomas. Classifying regular events in symbolic logic. *Journal of Computer and System Sciences*, 25(3):360–376, 1982.

[TW98] 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, 1998.