A generic polynomial time approach to separation by first-order logic without quantifier alternation

Thomas Place
LaBRI, Bordeaux University, France

Marc Zeitoun
LaBRI, Bordeaux University, France

Abstract

We look at classes of languages associated to the fragment of first-order logic $\mathcal{B}\Sigma_1$, in which quantifier alternations are disallowed. Each class is defined by choosing the set of predicates on positions that may be used. Two key such fragments are those equipped with the linear ordering and possibly the successor relation. Simon and Knast proved that these two variants have decidable *membership*: "does an input regular language belong to the class ?". We rely on a characterization of $\mathcal{B}\Sigma_1$ by the operator *BPol*: given an input class $\mathcal{C}$, it outputs a class $BPol(\mathcal{C})$ that corresponds to a variant of $\mathcal{B}\Sigma_1$ equipped with special predicates associated to $\mathcal{C}$. We extend the above results in two orthogonal directions. First, we use two kinds of inputs: classes $\mathcal{G}$ of *group languages* (*i.e.*, recognized by a DFA in which each letter induces a permutation of the states) and extensions thereof, written $\mathcal{G}^+$. The classes $BPol(\mathcal{G})$ and $BPol(\mathcal{G}^+)$ capture many natural variants of $\mathcal{B}\Sigma_1$ which use predicates such as the linear ordering, the successor, the modular predicates or the alphabetic modular predicates.

Second, instead of membership, we explore the more general separation problem: decide if two regular languages can be separated by a language from the class under study. We show that separation is decidable for $BPol(\mathcal{G})$ and $BPol(\mathcal{G}^+)$ when this is the case for $\mathcal{G}$. This was known for $BPol(\mathcal{G})$ and for two particular classes of the form $BPol(\mathcal{G}^+)$. Yet, the algorithms were indirect and relied on involved frameworks, yielding poor upper complexity bounds. In contrast, the approach of the paper is direct. We work only with elementary concepts (mainly, finite automata). Our main contribution consists in polynomial time Turing reductions from both $BPol(\mathcal{G})$- and $BPol(\mathcal{G}^+)$-separation to $\mathcal{G}$-separation. This yields polynomial algorithms for many key variants of $\mathcal{B}\Sigma_1$, including those equipped with the linear ordering and possibly the successor and/or the modular predicates.

2012 ACM Subject Classification Theory of computation → Formal languages and automata theory; Theory of computation → Regular languages

Keywords and phrases Automata, Separation, Covering, Concatenation hierarchies, Group languages

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2022.29

Related Version *Full version of the paper*: `https://arxiv.org/abs/2210.00946` [25]

Funding Supported by the DeLTA project (ANR-16-CE40-0007)

1 Introduction

An important question in automata theory is to precisely understand the prominent classes of regular languages of finite words. We are interested in the classes associated to a piece of syntax (such as regular expressions or logic), whose purpose is to specify the languages of such classes. In the paper, we formalize the goal of "understanding a given class $\mathcal{C}$" by looking at a decision problem: $\mathcal{C}$-separation. It takes two regular languages L_1, L_2 as input and asks whether there exists $K \in \mathcal{C}$ such that $L_1 \subseteq K$ and $K \cap L_2 = \emptyset$. The key idea is that obtaining an algorithm for $\mathcal{C}$-separation requires a solid understanding of $\mathcal{C}$.

We investigate a family of classes associated to a fragment of first-order logic written $\mathcal{B}\Sigma_1$. The sentences of $\mathcal{B}\Sigma_1$ are Boolean combinations of *existential* formulas, *i.e.*, whose prenex normal form has the shape $\exists x_1 \exists x_2 \cdots \exists x_k \varphi$, with φ quantifier-free. Several classes are associated to $\mathcal{B}\Sigma_1$, each determined by the predicates on positions that we allow. In the literature, standard examples of predicates include the linear order "$<$" [27], the successor relation "$+1$" [9] or modular predicates "MOD" [5]. Thus, a generic approach is desirable.

We tackle languages associated to $\mathcal{B}\Sigma_1$ through the operator $\mathcal{C} \mapsto BPol(\mathcal{C})$ defined on classes of languages. It is the composition of the polynomial closure $\mathcal{C} \mapsto Pol(\mathcal{C})$ and the Boolean closure $\mathcal{C} \mapsto Bool(\mathcal{C})$ operators: $BPol(\mathcal{C}) = Bool(Pol(\mathcal{C}))$. Recall that the polynomial closure of a class $\mathcal{C}$ consists of all finite unions of languages of the form $L_0 a_1 L_1 \cdots a_n L_n$, where $n \geq 0$, each a_i is a letter and each L_i belongs to $\mathcal{C}$. Indeed, many classes associated to $\mathcal{B}\Sigma_1$ are of the form $BPol(\mathcal{C})$ [34, 20]. In this paper, we look at specific input classes $\mathcal{C}$.

The *group languages* are those recognized by a finite group, or equivalently by a permutation automaton [33] (*i.e.*, which is complete, deterministic *and* co-deterministic). We consider input classes that are either a class $\mathcal{G}$ consisting of group languages, or a well-suited extension thereof, $\mathcal{G}^+$ (roughly, $\mathcal{G}^+$ is the least Boolean algebra containing $\mathcal{G}$ and the singleton $\{\varepsilon\}$). It is known [20] that if $\mathcal{G}$ is a class of group languages, then $BPol(\mathcal{G}) = \mathcal{B}\Sigma_1(<, \mathbb{P}_\mathcal{G})$ and $BPol(\mathcal{G}^+) = \mathcal{B}\Sigma_1(<, +1, \mathbb{P}_\mathcal{G})$. Here, $\mathbb{P}_\mathcal{G}$ is a set of predicates associated to $\mathcal{G}$: each language L in $\mathcal{G}$ gives rise to a predicate $P_L(x)$, which selects all positions x in a word w such that the prefix of w up to position x (excluded) belongs to L. This captures most of the natural examples. In particular, we get signatures including the aforementioned predicates, such as $\{<\}$, $\{<, +1\}$, $\{<, MOD\}$ and $\{<, +1, MOD\}$ (we provide some more examples in the paper).

State of the art. Historically, $BPol(\mathcal{G})$ and $BPol(\mathcal{G}^+)$ were first investigated for particular input classes. A prominent example is the class of piecewise testable languages [27], *i.e.*, the class $BPol(\mathrm{ST}) = \mathcal{B}\Sigma_1(<)$ where $\mathrm{ST} = \{\emptyset, A^*\}$. It was shown that $BPol(\mathrm{ST})$-separation is decidable in [1] using technical algebraic arguments. Simpler polynomial time algorithms were discovered later [17, 6]. There also exists an involved specialized separation algorithm [36] for $BPol(\mathrm{MOD}) = \mathcal{B}\Sigma_1(<, MOD)$, where MOD is the class of modulo languages. Decidability can be lifted to $BPol(\mathrm{ST}^+) = \mathcal{B}\Sigma_1(<, +1)$ (the languages of dot-depth one [9]) and to $BPol(\mathrm{MOD}^+) = \mathcal{B}\Sigma_1(<, +1, MOD)$ via transfer results [22, 16]. Unfortunately, this approach yields an exponential complexity blow-up. Recently, a generic approach was developed for $BPol(\mathcal{G})$. It is proved in [21] that if $\mathcal{G}$ is a class of group languages with mild hypotheses, $BPol(\mathcal{G})$-separation is decidable when $\mathcal{G}$-separation is decidable. Yet, this generic approach is indirect and considers a more general problem: *covering*. Because of this, the algorithms and their proofs are complex and rely on an intricate framework [19], yielding poor upper complexity bounds. This contrasts with the simple polynomial time procedures presented in [17, 6] for $BPol(\mathrm{ST})$. No generic result of this kind is known for the classes $BPol(\mathcal{G}^+)$.

Contributions. We give *generic polynomial time Turing reductions* from $BPol(\mathcal{G})$- and $BPol(\mathcal{G}^+)$-separation to $\mathcal{G}$-separation, where $\mathcal{G}$ is a class of group languages with mild prop-

erties. We present them as greatest fixpoint procedures which use an oracle for $\mathcal{G}$-separation at each step and run in *polynomial time* (for input languages represented by nondeterministic finite automata). While the proofs are involved, they are self-contained and based exclusively on elementary concepts from automata theory. No particular knowledge on group theory is required to follow them: we only use immediate consequences of the definition of a group.

For $BPol(\mathcal{G})$, this new approach is a significant improvement on the results of [21]. While we do reuse some ideas of [21], we complement them with new ones and the presentation is independent. We get a simpler algorithm, which requires only basic notions from automata theory. In particular, one direction of the proof describes a generic construction for building separators in $BPol(\mathcal{G})$ (when they exist). This serves our main objective: understanding classes of languages. In addition, we obtain much better complexity upper bounds on $BPol(\mathcal{G})$-separation. Finally, our techniques can handle $BPol(\mathcal{G}^+)$ as well. This was not the case in [21]: the generic reduction from $BPol(\mathcal{G}^+)$-separation to $\mathcal{G}$-separation is a new result.

These results apply to several key classes. Separation is decidable in polynomial time for $\mathrm{ST} = \{\emptyset, A^*\}$, for the class MOD of modulo languages and for the class GR of *all* group languages [26]. Hence, the problem is also decidable in polynomial time for $BPol(\mathrm{ST})$ (*i.e.*, $\mathcal{B}\Sigma_1(<)$), $BPol(\mathrm{ST}^+)$ (*i.e.*, $\mathcal{B}\Sigma_1(<, +1)$), $BPol(\mathrm{MOD})$ (*i.e.*, $\mathcal{B}\Sigma_1(<, MOD)$), $BPol(\mathrm{MOD}^+)$ (*i.e.*, $\mathcal{B}\Sigma_1(<, +1, MOD)$), $BPol(\mathrm{GR})$ and $BPol(\mathrm{GR}^+)$ (the logical characterization of the last two classes is not standard, yet they are quite prominent as well [11, 8]). This reproves a known result for $BPol(\mathrm{ST})$ (in fact, we essentially reprove the algorithm of [6]). The polynomial time upper bounds are new for all other classes. Another application is the class AMT of alphabet modulo testable languages (which are recognized by commutative groups): $BPol(\mathrm{AMT})$ and $BPol(\mathrm{AMT}^+)$ correspond to $\mathcal{B}\Sigma_1(<, AMOD)$ and $\mathcal{B}\Sigma_1(<, +1, AMOD)$ where "$AMOD$" is the set of *alphabetic modular predicates*. We obtain the decidability of separation for these classes (this is a new result for $BPol(\mathrm{AMT}^+)$). However, we do not get a polynomial time upper bound: this is because AMT-separation is co-NP-complete (see [26]).

Important remark. Eilenberg's theorem [7] connects some classes of regular languages (the "varieties of languages") with *varieties of finite monoids*. It raised the hope to solve decision problems on languages (such as membership) by translating them in terms of monoids and solving the resulting purely algebraic questions—without referring to languages anymore. In particular, Margolis and Pin [11, 13] characterized the algebraic counterpart of $BPol(\mathcal{G})$ in Eilenberg's correspondence (when $\mathcal{G}$ is a variety) as the "*semidirect product*" $\mathsf{J} * \mathsf{G}$, where J is the variety of monoids corresponding to $\mathcal{B}\Sigma_1(<)$ and G is the one corresponding to $\mathcal{G}$. The new purely algebraic question is then: "decide membership of a monoid in $\mathsf{J} * \mathsf{G}$". Tilson [35] developed an involved framework to reformulate membership in semidirect products in terms of categories, which was successfully exploited to handle $(\mathsf{J} * \mathsf{G})$-membership [8, 28].

Our results are completely independent from this algebraic approach. To clarify, we do use combinatorics on monoids. Yet, our motivations and techniques are disconnected from the theory of varieties of monoids, which is a distinct field. We avoid it by choice: while the above approach highlights an interesting connection between two fields, it is not necessarily desirable when looking back at our primary goal, understanding *classes of languages*. Indeed, a detour via varieties of monoids would obfuscate the intuition at the language level. Fortunately, this paper shows that this detour can be bypassed, while getting *stronger* results. First, our results are more general: they apply to *separation*, and not only membership. It is not clear at all that this can be obtained in the context of monoid varieties, as we rely strongly on the definition of $BPol$: we work with languages of the form $L_0 a_1 L_1 \cdots a_n L_n$, for $L_i \in \mathcal{G}$. Second, we can handle $BPol(\mathcal{G}^+)$, thus capturing the successor relation on the logical side. As far as we know, the only class of this kind captured by the above framework is $BPol(\mathrm{ST}^+)$ (these

are the well-known dot-depth one languages [30]). Third, using the above approach requires *varieties* of languages as input classes. This, for example, excludes the class $BPol(\mathrm{MOD})$. This does not mean that this class cannot be handled by algebraic techniques: this was actually done by Straubing [31, 15], who rebuilt the whole theory to be able to handle such classes. In contrast, our result applies *uniformly* to MOD.

Organization of the paper. We present the objects that we investigate and terminology in Section 2. We introduce separation and the techniques that we use to handle it in Section 3. Finally, we present our results for $BPol(\mathcal{G})$- and $BPol(\mathcal{G}^+)$-separation in Section 4. Due to space limitations, some proofs are only available in the full version of the paper [25].

2 Preliminaries

2.1 Words, regular languages and classes

We fix a finite *alphabet* A for the paper. As usual, A^* denotes the set of all finite words over A, including the empty word ε. We let $A^+ = A^* \setminus \{\varepsilon\}$. For $u, v \in A^*$, we let uv be the word obtained by concatenating u and v. A *language* is a subset of A^*. We denote the singleton language $\{u\}$ by u. We lift concatenation to languages: for $K, L \subseteq A^*$, we let $KL = \{uv \mid u \in K \text{ and } v \in L\}$. We shall consider *marked products*: given languages $L_0, \ldots, L_n \subseteq A^*$, a marked product of $L_0, \ldots, L_n$ is a product of the form $L_0 a_1 L_1 \cdots a_n L_n$ where $a_1, \ldots, a_n \in A$ (note that "L_0" is a marked product: this is the case $n = 0$).

Regular languages. In the paper, we consider *regular* languages. A nondeterministic finite automaton (NFA) is a pair $\mathcal{A} = (Q, \delta)$ where Q is a finite set of states, and $\delta \subseteq Q \times A \times Q$ is a set of transitions. We now define the languages recognized by $\mathcal{A}$. Given $q, r \in Q$ and $w \in A^*$, we say that there exists a *run labeled by w from q to r* (in $\mathcal{A}$) if there exist $q_0, \ldots, q_n \in Q$ and $a_1, \ldots, a_n \in A$ such that $w = a_1 \cdots a_n$, $q_0 = q$, $q_n = r$ and $(q_{i-1}, a_i, q_i) \in \delta$ for every $1 \leq i \leq n$. Given two sets $I, F \subseteq Q$, we write $L_{\mathcal{A}}(I, F) \subseteq A^*$ for the language of all words $w \in A^*$ such that there exist $q \in I$, $r \in F$, and a run labeled by w from q to r in $\mathcal{A}$. We say that a language $L \subseteq A^*$ is *recognized* by $\mathcal{A}$ if and only if there exist $I, F \subseteq Q$ such that $L = L_{\mathcal{A}}(I, F)$. The regular languages are those which can be recognized by an NFA.

We also use NFAs with *ε-transitions*. In such an NFA $\mathcal{A} = (Q, \delta)$, a transition may also be labeled by the empty word "ε" (that is, $\delta \subseteq Q \times (A \cup \{\varepsilon\}) \times Q$). We use the standard semantics: an ε-transition can be taken without consuming an input letter. Note that unless otherwise specified, the NFAs that we consider are assumed to be *without* ε-transitions.

Classes. A *class* of languages is a set of languages. A *lattice* is a class containing $\emptyset$ and A^* and closed under both union and intersection. Moreover, a *Boolean algebra* is a lattice closed under complement. Finally, a class $\mathcal{C}$ is *quotient-closed* when for all $L \in \mathcal{C}$ and all $v \in A^*$, the languages $v^{-1}L = \{w \in A^* \mid vw \in L\}$ and $Lv^{-1} = \{w \in A^* \mid wv \in L\}$ both belong to $\mathcal{C}$ as well. A *positive prevariety* (resp. a *prevariety*) is a quotient-closed lattice (resp. a quotient-closed Boolean algebra) containing *regular languages only*.

Group languages. A *monoid* is a set M equipped with a multiplication $s, t \mapsto st$, which is associative and has a neutral element denoted by "1_M". Observe that A^* endowed with concatenation is a monoid (ε is the neutral element). It is well-known that a language L is regular if and only if it is *recognized* by a morphism $\alpha : A^* \to M$ into a *finite* monoid M, *i.e.*, there exists $F \subseteq M$ such that $L = \alpha^{-1}(F)$. We now restrict this definition: a monoid G is a *group* if every element $g \in G$ has an inverse $g^{-1} \in G$, *i.e.*, such that $gg^{-1} = g^{-1}g = 1_G$. A "*group language*" is a language recognized by a morphism into a *finite group*.

We consider classes $\mathcal{G}$ that are group prevarieties (*i.e.*, containing group languages only). We let GR be the class of *all* group languages. Another important example is the class AMT of *alphabet modulo testable languages*. For every $w \in A^*$ and every $a \in A$, we write $\#_a(w) \in \mathbb{N}$ for the number of occurrences of "a" in w. The class AMT consists in all finite Boolean combinations of languages $\{w \in A^* \mid \#_a(w) \equiv k \bmod m\}$ where $a \in A$ and $k, m \in \mathbb{N}$ are such that $k < m$. One may verify that these are exactly the languages recognized by commutative groups. We also consider the class MOD, which consists in all finite Boolean combinations of languages $\{w \in A^* \mid |w| \equiv k \bmod m\}$ with $k, m \in \mathbb{N}$ such that $k < m$. Finally, we write ST for the trivial class $\mathrm{ST} = \{\emptyset, A^*\}$. One may verify that GR, AMT, MOD and ST are all group prevarieties.

One may verify that $\{\varepsilon\}$ and A^+ are *not* group languages. This motivates the next definition: the *well-suited extension of a class* $\mathcal{C}$, denoted by $\mathcal{C}^+$, consists of all languages of the form $L \cap A^+$ or $L \cup \{\varepsilon\}$ where $L \in \mathcal{C}$. The next lemma follows from the definition.

▶ **Lemma 1.** *Let $\mathcal{C}$ be a prevariety. Then, $\mathcal{C}^+$ is a prevariety containing $\{\varepsilon\}$ and A^+.*

2.2 Polynomial and Boolean closure

We investigate two operators that one may apply to a class $\mathcal{C}$. The *Boolean closure* of $\mathcal{C}$, written $\mathit{Bool}(\mathcal{C})$, is the least Boolean algebra containing $\mathcal{C}$. The *polynomial closure* of $\mathcal{C}$, denoted by $\mathit{Pol}(\mathcal{C})$, consists of all finite unions of marked products $L_0 a_1 L_1 \cdots a_n L_n$ where $L_0, \ldots, L_n \in \mathcal{C}$ and $a_1, \ldots, a_n \in A$. Finally, we write $\mathit{BPol}(\mathcal{C})$ for $\mathit{Bool}(\mathit{Pol}(\mathcal{C}))$. If $\mathcal{C}$ is a prevariety, then $\mathit{Pol}(\mathcal{C})$ is a positive prevariety and $\mathit{BPol}(\mathcal{C})$ is a prevariety. Proving that $\mathit{Pol}(\mathcal{C})$ is closed under intersection is not immediate. It was shown by Arfi [2] (see also [14, 20]).

▶ **Theorem 2.** *If $\mathcal{C}$ is a prevariety, $\mathit{Pol}(\mathcal{C})$ is a positive prevariety and $\mathit{BPol}(\mathcal{C})$ is a prevariety.*

The two operators *Pol* and *Bool* induce standard classifications called concatenation hierarchies: for a prevariety $\mathcal{C}$, the *concatenation hierarchy of basis* $\mathcal{C}$ is built from $\mathcal{C}$ by alternatively applying the operators *Pol* and *Bool*. We are interested in $\mathit{BPol}(\mathcal{C})$, which is level *one* in the concatenation hierarchy of basis $\mathcal{C}$. We look at bases that are either a group prevariety $\mathcal{G}$ or its well-suited extension $\mathcal{G}^+$. Most of the prominent concatenation hierarchies in the literature use such bases. This is in part motivated by the logical characterization of concatenation hierarchies, due to Thomas [34]. We briefly recall it for the level one.

Consider a word $w = a_1 \cdots a_{|w|} \in A^*$. We view w as a linearly ordered set of $|w| + 2$ positions $\{0, 1, \ldots, |w|, |w|+1\}$ such that each position $1 \leq i \leq |w|$ carries the label $a_i \in A$ (on the other hand, 0 and $|w|+1$ are artificial unlabeled leftmost and rightmost positions). We use first-order logic to describe properties of words: a sentence can quantify over the positions of a word and use a predetermined set of predicates to test properties of these positions. We also allow two constants "min" and "max" interpreted as the artificial unlabeled positions 0 and $|w|+1$ in a given word w. A first-order sentence φ defines the language of all words satisfying the property stated by φ. We use several kinds of predicates. For each $a \in A$, we associate a unary predicate (also denoted by a), which selects the positions labeled by "a". We also use two binary predicates: the (strict) linear order "$<$" and the successor relation "$+1$". Finally, we associate a set of predicates $\mathbb{P}_{\mathcal{G}}$ to each group prevariety $\mathcal{G}$. Every $L \in \mathcal{G}$ yields a unary predicate P_L in $\mathbb{P}_{\mathcal{G}}$, which is interpreted as follows. Let $w = a_1 \cdots a_{|w|} \in A^*$. The unary predicate P_L selects all positions $i \in \{0, \ldots, |w| + 1\}$ such that $i \neq 0$ and $a_1 \cdots a_{i-1} \in L$.

▶ **Example 3.** The sentence "$\exists x \exists y\ (x < y) \wedge a(x) \wedge b(y)$" defines the language $A^* a A^* b A^*$. The sentence "$\exists x \exists y\ a(x) \wedge c(y) \wedge (y + 1 = max)$" defines $A^* a A^* c$. Finally, if $L = (AA)^* \in \mathrm{MOD}$ (the words of even length), the sentence "$\exists x\ a(x) \wedge P_L(x)$" defines the language $(AA)^* a A^*$.

The fragment of first-order logic containing exactly the Boolean combinations of existential first-order sentences is denoted by "$\mathcal{B}\Sigma_1$". Let $\mathcal{G}$ be a group prevariety. We write $\mathcal{B}\Sigma_1(<, \mathbb{P}_\mathcal{G})$ for the class of all languages defined by a sentence of $\mathcal{B}\Sigma_1$ using only the label predicates, the linear order "$<$" and those in $\mathbb{P}_\mathcal{G}$. Moreover, we write $\mathcal{B}\Sigma_1(<, +1, \mathbb{P}_\mathcal{G})$ for the class of all languages defined by a sentence of $\mathcal{B}\Sigma_1$, which additionally allows the successor predicate "$+1$". The following proposition follows from the results of [20, 24].

▶ **Proposition 4.** *Let $\mathcal{G}$ be a group prevariety. We have $BPol(\mathcal{G}) = \mathcal{B}\Sigma_1(<, \mathbb{P}_\mathcal{G})$ and $BPol(\mathcal{G}^+) = \mathcal{B}\Sigma_1(<, +1, \mathbb{P}_\mathcal{G})$.*

Key examples. The basis $\mathrm{ST} = \{\emptyset, A^*\}$ yields the *Straubing-Thérien hierarchy* [29, 32] (hence the notation of this basis). Its level one is the class of piecewise testable languages [27]. Its well-suited extension ST^+ induces the *dot-depth hierarchy* [3]. In particular, $BPol(\mathrm{ST})$ and $BPol(\mathrm{ST}^+)$ correspond to $\mathcal{B}\Sigma_1(<)$ and $\mathcal{B}\Sigma_1(<, +1)$, as all predicates in $\mathbb{P}_\mathrm{ST}$ are trivial. The hierarchies of bases MOD and MOD^+ are also prominent (see for example [5, 10, 36]). The classes $BPol(\mathrm{MOD})$ and $BPol(\mathrm{MOD}^+)$ correspond to $\mathcal{B}\Sigma_1(<, MOD)$ and $\mathcal{B}\Sigma_1(<, +1, MOD)$ where "MOD" is the set of *modular predicates* (for all $r, q \in \mathbb{N}$ such that $r < q$, it contains a unary predicate $M_{r,q}$ selecting the positions i such that $i \equiv r \bmod q$). Similarly, $BPol(\mathrm{AMT})$ and $BPol(\mathrm{AMT}^+)$ correspond to $\mathcal{B}\Sigma_1(<, AMOD)$ and $\mathcal{B}\Sigma_1(<, +1, AMOD)$ where "$AMOD$" is the set of *alphabetic modular predicates* (for all $a \in A$ and $r, q \in \mathbb{N}$ such that $r < q$, it contains a unary predicate $M^a_{r,q}$ selecting the positions i such the that number of positions $j < i$ with label a is congruent to r modulo q). Finally, the group hierarchy, whose basis is GR is also prominent [11, 8], though its logical characterization is not standard.

Properties. We present a key ingredient [23, Lemma 3.6]. It describes a concatenation principle for the classes $BPol(\mathcal{C})$ based on the notion of "cover". Given a language L, a cover of L is a *finite* set $\mathbf{K}$ of languages satisfying $L \subseteq \bigcup_{K \in \mathbf{K}} K$. If $\mathcal{D}$ is a class, a $\mathcal{D}$-cover of L is a cover $\mathbf{K}$ of L such that $\mathbf{K} \subseteq \mathcal{D}$.

▶ **Proposition 5.** *Let $\mathcal{C}$ be a prevariety, $n \in \mathbb{N}$, $L_0, \ldots, L_n \in Pol(\mathcal{C})$ and $a_1, \ldots, a_n \in A$. If $\mathbf{H}_i$ is a $BPol(\mathcal{C})$-cover of L_i for all $i \leq n$, then there is a $BPol(\mathcal{C})$-cover $\mathbf{K}$ of $L_0 a_1 L_1 \cdots a_n L_n$ such that for all $K \in \mathbf{K}$, there exists $H_i \in \mathbf{H}_i$ for each $i \leq n$ satisfying $K \subseteq H_0 a_1 H_1 \cdots a_n H_n$.*

For applying Proposition 5, we need a language $L_0 a_1 L_1 \cdots a_n L_n$ with $L_0, \ldots, L_n \in Pol(\mathcal{C})$. The next tailored statements build such languages when $\mathcal{C} = \mathcal{G}$ or $\mathcal{G}^+$ for a group prevariety $\mathcal{G}$. While simple, these results are central: this is the unique place where we use the fact that $\mathcal{G}$ contains only *group languages*. Let $L \subseteq A^*$. With every word $w = a_1 \cdots a_n \in A^*$, we associate the language $\uparrow_L w = L a_1 L \cdots a_n L \subseteq A^*$ (we let $\uparrow_L \varepsilon = L$). We first present the statement for the case $\mathcal{C} = \mathcal{G}$, which can also be found in [4, Prop. 3.11].

▶ **Proposition 6.** *Let $H \subseteq A^*$ be a language and $L \subseteq A^*$ be a group language containing ε. There exists a cover $\mathbf{K}$ of H such that every $K \in \mathbf{K}$ is of the form $K = \uparrow_L w$ for some $w \in H$.*

The next statement, useful for the case $\mathcal{C} = \mathcal{G}^+$, is a corollary of Proposition 6. Let $\mathcal{A} = (Q, \delta)$ be an NFA. Moreover, let $w, z \in A^*$. We say that z is a *left $\mathcal{A}$-loop* for w if for every $q, r \in Q$ such that $w \in L_\mathcal{A}(q, r)$, there exists $s \in Q$ such that $z \in L_\mathcal{A}(q, s) \cap L_\mathcal{A}(s, s)$ and $zw \in L_\mathcal{A}(s, r)$ (in particular, $zz^*zw \subseteq L_\mathcal{A}(q, r)$). Symmetrically, we say that z is a *right $\mathcal{A}$-loop* for w if for every $q, r \in Q$ such that $w \in L_\mathcal{A}(q, r)$, there exists $s \in Q$ such that $wz \in L_\mathcal{A}(q, s)$ and $z \in L_\mathcal{A}(s, s) \cap L_\mathcal{A}(s, r)$ (in particular, $wzz^*z \subseteq L_\mathcal{A}(q, r)$).

Now, given an arbitrary word $w \in A^*$, an *$\mathcal{A}$-guarded decomposition of w* is a tuple $(w_1, \ldots, w_{n+1})$ for some $n \in \mathbb{N}$ where $w_1 \in A^*$ and $w_i \in A^+$ for $2 \leq i \leq n+1$, and such that $w = w_1 \cdots w_{n+1}$ and, if $n \geq 1$, then for every i satisfying $1 \leq i \leq n$, there exists a *nonempty* word $z_i \in A^+$ which is a right $\mathcal{A}$-loop for w_i and a left $\mathcal{A}$-loop for w_{i+1}.

▶ **Proposition 7.** *Let $H \subseteq A^*$ be a language, $\mathcal{A}$ be an NFA and $L \subseteq A^*$ be a group language containing ε. There exists a cover $\mathbf{K}$ of H such that for each $K \in \mathbf{K}$, there exist a word $w \in H$ and an $\mathcal{A}$-guarded decomposition $(w_1, \ldots, w_{n+1})$ of w for some $n \in \mathbb{N}$ such that $K = w_1 L \cdots w_n L w_{n+1}$ (if $n = 0$, then $K = \{w_1\}$).*

3 Separation framework

In order to investigate a given class $\mathcal{C}$, we rely on a generic decision problem that one may associate to it: *$\mathcal{C}$-separation*. We first define it and then present a variant, "tuple separation", that we shall require as a proof ingredient.

3.1 The separation problem

Consider two languages $L_0, L_1 \subseteq A^*$. We say that a third language $K \subseteq A^*$ *separates* L_0 from L_1 when $L_0 \subseteq K$ and $K \cap L_1 = \emptyset$. Then, given an arbitrary class $\mathcal{C}$, we say that L_0 is *$\mathcal{C}$-separable* from L_1 when there exists $K \in \mathcal{C}$ that separates L_0 from L_1. For every class $\mathcal{C}$, the *$\mathcal{C}$-separation problem* takes *two* regular languages L_0 and L_1 as input (in the paper, they are represented by NFAs) and asks whether L_0 is $\mathcal{C}$-separable from L_1. We complete the definition with a useful result, which holds when $\mathcal{C}$ is a positive prevariety.

▶ **Lemma 8.** *Let $\mathcal{C}$ be a positive prevariety and $L_0, L_1, H_0, H_1 \subseteq A^*$. If L_0 is not $\mathcal{C}$-separable from L_1 and H_0 is not $\mathcal{C}$-separable from H_1 then $L_0 H_0$ is not $\mathcal{C}$-separable from $L_1 H_1$.*

In the paper, we look at $\mathcal{C}$-separation when $\mathcal{C} = BPol(\mathcal{G})$ or $BPol(\mathcal{G}^+)$ for a group prevariety $\mathcal{G}$. We prove that in these two cases, there are polynomial time (Turing) reductions to $\mathcal{G}$-separation. We now introduce terminology that we shall use to present the algorithms.

Framework. Consider a class $\mathcal{C}$ and an NFA $\mathcal{A} = (Q, \delta)$. We associate a set $\mathcal{I}_{\mathcal{C}}[\mathcal{A}] \subseteq Q^4$: the *inseparable $\mathcal{C}$-quadruples* associated to $\mathcal{A}$. We define,

$$\mathcal{I}_{\mathcal{C}}[\mathcal{A}] = \big\{(q, r, s, t) \in Q^4 \mid L_{\mathcal{A}}(q, r) \text{ is } \underline{\text{not}} \text{ } \mathcal{C}\text{-separable from } L_{\mathcal{A}}(s, t)\big\}.$$

The next easy result connects $\mathcal{C}$-separation to this set, for input languages given by NFAs.

▶ **Proposition 9.** *Let $\mathcal{C}$ be a lattice. Consider an NFA $\mathcal{A} = (Q, \delta)$ and four sets of states $I_1, F_1, I_2, F_2 \subseteq Q$. The two following conditions are equivalent:*

1. *$L_{\mathcal{A}}(I_1, F_1)$ is $\mathcal{C}$-separable from $L_{\mathcal{A}}(I_2, F_2)$.*
2. *$(I_1 \times F_1 \times I_2 \times F_2) \cap \mathcal{I}_{\mathcal{C}}[\mathcal{A}] = \emptyset$.*

Clearly, given as input two regular languages recognized by NFAs, one may compute in polynomial time a single NFA recognizing both languages. Hence, Proposition 9 yields a polynomial time reduction from $\mathcal{C}$-separation to the problem of computing $\mathcal{I}_{\mathcal{C}}[\mathcal{A}]$ from an input NFA. Naturally, this does not necessarily mean that there exists a polynomial time algorithm for $\mathcal{C}$-separation: depending on $\mathcal{C}$, computing $\mathcal{I}_{\mathcal{C}}[\mathcal{A}]$ may or may not be costly.

We introduce a key definition for manipulating $\mathcal{I}_{\mathcal{C}}[\mathcal{A}]$, for an NFA $\mathcal{A} = (Q, \delta)$. Let $S \subseteq Q^4$ and $\mathbf{K}$ be a finite set of languages. We say that $\mathbf{K}$ *is separating for S* when for every $(q, r, s, t) \in Q^4$ and every $K \in \mathbf{K}$, if K intersects both $L_{\mathcal{A}}(q, r)$ and $L_{\mathcal{A}}(s, t)$, then $(q, r, s, t) \in S$. Then, $\mathcal{I}_{\mathcal{C}}[\mathcal{A}]$ is the smallest set of 4-tuples admitting a $\mathcal{C}$-cover of A^* which is separating for it.

▶ **Lemma 10.** *Let $\mathcal{C}$ be a Boolean algebra and $\mathcal{A} = (Q, \delta)$ be an NFA. Then the following holds:*

- *There exists a $\mathcal{C}$-cover $\mathbf{K}$ of A^* which is separating for $\mathcal{I}_{\mathcal{C}}[\mathcal{A}]$.*
- *Let $S \subseteq Q^4$. If there exists a $\mathcal{C}$-cover $\mathbf{K}$ of A^* which is separating for S, then $\mathcal{I}_{\mathcal{C}}[\mathcal{A}] \subseteq S$.*

Controlled separation. We present additional terminology tailored to the classes built from a group prevariety. Consider two classes $\mathcal{C}$ and $\mathcal{D}$ (in practice, $\mathcal{D}$ will be a group prevariety $\mathcal{G}$ and $\mathcal{C}$ will be either $BPol(\mathcal{G})$ or $BPol(\mathcal{G}^+)$). Let $L_0, L_1 \subseteq A^*$. We say that L_0 is $\mathcal{C}$-separable from L_1 *under* $\mathcal{D}$*-control* if there exists $H \in \mathcal{D}$ such that $\varepsilon \in H$ and $L_0 \cap H$ is $\mathcal{C}$-separable from $L_1 \cap H$. Given an NFA $\mathcal{A} = (Q, \delta)$, we associate a set $\mathcal{I}_{\mathcal{C}}[\mathcal{D}, \mathcal{A}] \subseteq Q^4$:

$$\mathcal{I}_{\mathcal{C}}[\mathcal{D}, \mathcal{A}] = \big\{(q, r, s, t) \in Q^4 \mid L_{\mathcal{A}}(q, r) \text{ is } \underline{\text{not}} \text{ } \mathcal{C}\text{-separable from } L_{\mathcal{A}}(s, t) \text{ under } \mathcal{D}\text{-control}\big\}.$$

Clearly, we have $\mathcal{I}_{\mathcal{C}}[\mathcal{D}, \mathcal{A}] \subseteq \mathcal{I}_{\mathcal{C}}[\mathcal{A}]$. Let us connect this new definition to the notion of separating cover presented above. In this case as well, this will be useful in proof arguments.

▶ **Lemma 11.** *Let $\mathcal{C}$ and $\mathcal{D}$ be Boolean algebras such that $\mathcal{D} \subseteq \mathcal{C}$ and let $\mathcal{A} = (Q, \delta)$ be an NFA. The following properties hold:*

- *There exists $L \in \mathcal{D}$ with $\varepsilon \in L$, and a $\mathcal{C}$-cover $\mathbf{K}$ of L which is separating for $\mathcal{I}_{\mathcal{C}}[\mathcal{D}, \mathcal{A}]$.*
- *Let $S \subseteq Q^4$. If there exist $L \in \mathcal{D}$ with $\varepsilon \in L$, and a $\mathcal{C}$-cover $\mathbf{K}$ of L which is separating for S, then $\mathcal{I}_{\mathcal{C}}[\mathcal{D}, \mathcal{A}] \subseteq S$.*

This notion is only useful if $\{\varepsilon\} \notin \mathcal{D}$. If $\{\varepsilon\} \in \mathcal{D}$, then L_0 is $\mathcal{C}$-separable from L_1 under $\mathcal{D}$-control if and only if either $\varepsilon \notin L_0$ or $\varepsilon \notin L_1$. This is why the notion is designed for group prevarieties: if $\mathcal{G}$ is such a class, then $\{\varepsilon\} \notin \mathcal{G}$. In this case, if $\mathcal{C} \in \{\mathcal{G}, \mathcal{G}^+\}$, then the set $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{G}, \mathcal{A}]$ carries more information than $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}]$. This is useful for the computation: rather than computing $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}]$ directly, our procedures first compute $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{G}, \mathcal{A}]$. The proof is based on Propositions 5 and 6 (the latter requires $\mathcal{G}$ to consist of group languages).

▶ **Proposition 12.** *Let $\mathcal{G}$ be a group prevariety, let $\mathcal{C}$ be a prevariety such that $\mathcal{G} \subseteq \mathcal{C}$ and let $\mathcal{A} = (Q, \delta)$ be an NFA. Then, $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}]$ is the least set $S \subseteq Q^4$ that contains $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{G}, \mathcal{A}]$ and satisfies the two following conditions:*

1. *For all $q, r, s, t \in Q$ and $a \in A$, if $(q, a, r), (s, a, t) \in \delta$, then $(q, r, s, t) \in S$.*
2. *For all $(q_1, r_1, s_1, t_1), (q_2, r_2, s_2, t_2) \in S$, if $r_1 = q_2$ and $t_1 = s_2$, then $(q_1, r_2, s_1, t_2) \in S$.*

Proof. Let $S \subseteq Q^4$ be the least set containing $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{G}, \mathcal{A}]$ and satisfying both conditions. We prove that $S = \mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}]$. For $S \subseteq \mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}]$, since $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{G}, \mathcal{A}] \subseteq \mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}]$ by definition, it suffices to prove that $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}]$ satisfies both conditions in the proposition. First, consider $a \in A$ and $q, r, s, t \in Q$ such that $(q, a, r), (s, a, t) \in \delta$. We have $a \in L_{\mathcal{A}}(q, r)$ and $a \in L_{\mathcal{A}}(s, t)$. Hence, they are not $BPol(\mathcal{C})$-separable and $(q, r, s, t) \in \mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}]$. Now, let $(q_1, r_1, s_1, t_1), (q_2, r_2, s_2, t_2) \in \mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}]$ such that $r_1 = q_2$ and $t_1 = s_2$. For $i \in \{1, 2\}$, we know that $L_{\mathcal{A}}(q_i, r_i)$ is not $BPol(\mathcal{C})$-separable from $L_{\mathcal{A}}(s_i, t_i)$. Since $BPol(\mathcal{C})$ is a prevariety by Theorem 2, it follows from Lemma 8 that $L_{\mathcal{A}}(q_1, r_1)L_{\mathcal{A}}(q_2, r_2)$ is not $BPol(\mathcal{C})$ separable from $L_{\mathcal{A}}(s_1, t_1)L_{\mathcal{A}}(s_2, t_2)$. Since $r_1 = q_2$ and $t_1 = s_2$, it is immediate that $L_{\mathcal{A}}(q_1, r_1)L_{\mathcal{A}}(q_2, r_2) \subseteq L_{\mathcal{A}}(q_1, r_2)$ and $L_{\mathcal{A}}(s_1, t_1)L_{\mathcal{A}}(s_2, t_2) \subseteq L_{\mathcal{A}}(s_1, t_2)$. Hence, $L_{\mathcal{A}}(q_1, r_2)$ is not $BPol(\mathcal{C})$-separable from $L_{\mathcal{A}}(s_1, t_2)$ and we get $(q_1, r_2, s_1, t_2) \in \mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}]$ as desired.

We turn to the inclusion $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}] \subseteq S$. By Lemma 11, there exists $L \in \mathcal{G}$ such that $\varepsilon \in L$ and a $BPol(\mathcal{C})$-cover $\mathbf{V}$ of L which is separating for $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{G}, \mathcal{A}]$. By hypothesis, L is a group language and $\varepsilon \in L$. Hence, Proposition 6 yields a cover $\mathbf{P}$ of A^* such that every $P \in \mathbf{P}$ is of the form $P = \uparrow_L w_P$ for some word $w_P \in A^*$. Let $P \in \mathbf{P}$ and $a_1, \ldots, a_n \in A$ be the letters such that $w_P = a_1 \cdots a_n$. We have $P = La_1L \cdots a_nL$ by definition (if $w_P = \varepsilon$, then $P = L$). By definition, $L \in \mathcal{G} \subseteq Pol(\mathcal{C})$. Hence, since $\mathbf{V}$ is a $BPol(\mathcal{C})$-cover of L, Proposition 5 yields a $BPol(\mathcal{C})$-cover $\mathbf{K}_P$ of P such that for every $K \in \mathbf{K}_P$, there are $V_0, \ldots, V_n \in \mathbf{V}$ such that $K \subseteq V_0a_1V_1 \cdots a_nV_n$. We let $\mathbf{K} = \bigcup_{P \in \mathbf{P}} \mathbf{K}_P$. Since $\mathbf{P}$ is a cover of A^* and $\mathbf{K}_P$ is a $BPol(\mathcal{C})$-cover of P for each $P \in \mathbf{P}$, $\mathbf{K}$ is a $BPol(\mathcal{C})$-cover of A^*. We show that $\mathbf{K}$ is separating for S which implies that $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}] \subseteq S$ by Lemma 10.

Let $(q,r,s,t) \in Q^4$ and $K \in \mathbf{K}$ such that we have $x \in K \cap L_{\mathcal{A}}(q,r)$ and $y \in K \cap L_{\mathcal{A}}(s,t)$. We show that $(q,r,s,t) \in S$. We have $K \in \mathbf{K}_P$ for some $P \in \mathbf{P}$. Let $a_1, \ldots, a_n \in A$ such that $w_P = a_1 \cdots a_n$. By definition, there are $V_0, \ldots, V_n \in \mathbf{V}$ such that $K \subseteq V_0 a_1 V_1 \cdots a_n V_n$. Since $x, y \in K$, we get $x_i, y_i \in V_i$ for $0 \leq i \leq n$ such that $x = x_0 a_1 x_1 \cdots a_n x_n$ and $y = y_0 a_1 y_1 \cdots a_n y_n$. Since $x \in L_{\mathcal{A}}(q,r)$, we get $q_i, r_i \in Q$ for $0 \leq i \leq n$ such that $q_0 = q$, $r_n = r$, $x_i \in L_{\mathcal{A}}(q_i, r_i)$ for $0 \leq i \leq n$ and $(r_{i-1}, a_i, q_i) \in \delta$ for $1 \leq i \leq n$. Finally, since $y \in L_{\mathcal{A}}(s,t)$, we get $s_i, t_i \in Q$ for $0 \leq i \leq n$ such that $s_0 = s$, $t_n = t$, $y_i \in L_{\mathcal{A}}(s_i, t_i)$ for $0 \leq i \leq n$ and $(t_{i-1}, a_i, s_i) \in \delta$ for $1 \leq i \leq n$. Since S satisfies Condition 1 in the proposition, we get $(r_{i-1}, q_i, t_{i-1}, s_i) \in S$ for $1 \leq i \leq n$. Since $V_i \in \mathbf{V}$ which is separating for $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{G}, \mathcal{A}]$ and $x_i, y_i \in V_i$, we also get $(q_i, r_i, q_i, t_i) \in \mathcal{I}_{BPol(\mathcal{C})}[\mathcal{G}, \mathcal{A}]$ for $0 \leq i \leq n$. Thus, Condition 2 in the proposition yields $(q_0, r_0, s_n, t_n) \in S$, *i.e.* $(q,r,s,t) \in S$ as desired. ◀

Proposition 12 provides a least fixpoint algorithm for computing the set $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{A}]$ from $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{G}, \mathcal{A}]$. Combined with Proposition 9, this yields a polynomial time reduction from $BPol(\mathcal{C})$-separation to computing $\mathcal{I}_{BPol(\mathcal{C})}[\mathcal{G}, \mathcal{A}]$ from an NFA. We shall prove that when $\mathcal{C} \in \{\mathcal{G}, \mathcal{G}^+\}$, there are polynomial time reductions of the latter problem to $\mathcal{G}$-separation.

3.2 Tuple separation

This generalized variant of separation is taken from [18]. We shall use it as a proof ingredient: for every lattice $\mathcal{C}$, it is connected to the classical separation problem for $Bool(\mathcal{C})$. For every $n \geq 1$, we call "*n-tuple*" a tuple of n languages $(L_1, \ldots, L_n)$. In the sequel, given another language K, we shall write $(L_1, \ldots, L_n) \cap K$ for the n-tuple $(L_1 \cap K, \ldots, L_n \cap K)$. Let $\mathcal{C}$ be a lattice, we use induction on n to define the *$\mathcal{C}$-separable n-tuples*:

- If $n = 1$, a 1-tuple (L_1) is $\mathcal{C}$-separable when $L_1 = \emptyset$.
- If $n \geq 2$, an n-tuple $(L_1, \ldots, L_n)$ is $\mathcal{C}$-separable when there exists $K \in \mathcal{C}$ such that $L_1 \subseteq K$ and $(L_2, \ldots, L_n) \cap K$ is $\mathcal{C}$-separable. We call K a *separator* of $(L_1, \ldots, L_n)$.

One may verify that classical separation is the special case $n = 2$. We generalize $\mathcal{D}$-controlled separation to this setting. For a class $\mathcal{D}$, we say that an n-tuple $(L_1, \ldots, L_n)$ is $\mathcal{C}$-separable under $\mathcal{D}$-control if there exists $H \in \mathcal{D}$ such that $\varepsilon \in H$ and $(L_1, \ldots, L_n) \cap H$ is $\mathcal{C}$-separable.

We complete the definition with two simple properties of tuple separation. The second one is based on closure under quotients and generalizes Lemma 8.

▶ **Lemma 13.** *Let $\mathcal{C}$ be a lattice and let $(L_1, \ldots, L_n), (H_1, \ldots, H_n)$ be two n-tuples. If $L_1 \cap \cdots \cap L_n \neq \emptyset$, then $(L_1, \ldots, L_n)$ is not $\mathcal{C}$-separable. Moreover, if $L_i \subseteq H_i$ for every $i \leq n$ and $(L_1, \ldots, L_n)$ is not $\mathcal{C}$-separable, then $(H_1, \ldots, H_n)$ is not $\mathcal{C}$-separable either.*

▶ **Lemma 14.** *Let $\mathcal{C}$ be a positive prevariety, $n \geq 1$ and let $(L_1, \ldots, L_n), (H_1, \ldots, H_n)$ be two n-tuples, which are not $\mathcal{C}$-separable. Then, $(L_1 H_1, \ldots, L_n H_n)$ is not $\mathcal{C}$-separable either.*

A theorem of [18] connects tuple $\mathcal{C}$-separation for a lattice $\mathcal{C}$ to $Bool(\mathcal{C})$-separation: L_0 is $Bool(\mathcal{C})$-separable from L_1 if and only if $(L_0, L_1)^p$ is $\mathcal{C}$-separable for some $p \geq 1$. Here, $(L_0, L_1)^p$ denotes the $2p$-tuple obtained by concatenating p copies of (L_0, L_1). For example, $(L_0, L_1)^3 = (L_0, L_1, L_0, L_1, L_0, L_1)$. We use a corollary applying to $\mathcal{D}$-controlled separation.

▶ **Corollary 15.** *Let $\mathcal{C}$ and $\mathcal{D}$ be two lattices such that $\mathcal{D} \subseteq \mathcal{C}$ and let $L_0, L_1 \subseteq A^*$. The following properties are equivalent:*

1. *L_0 is $Bool(\mathcal{C})$-separable from L_1 under $\mathcal{D}$-control.*
2. *There exists $p \geq 1$ such that $(L_0, L_1)^p$ is $\mathcal{C}$-separable under $\mathcal{D}$-control.*

We only use the contrapositive of 1) ⇒ 2) in Corollary 15. We complete the presentation with two important lemmas about tuple separation for $Pol(\mathcal{D})$ and $Pol(\mathcal{D}^+)$. We use them to prove that tuples are not separable. Note that in practice, $\mathcal{D}$ will be a group prevariety $\mathcal{G}$. Yet, the results are true regardless of this hypothesis.

▶ **Lemma 16.** *Let $\mathcal{D}$ be a prevariety and $(L_1, \dots, L_n)$ an n-tuple which is not $Pol(\mathcal{D})$-separable under $\mathcal{D}$-control. Then, $(\{\varepsilon\}, L_1, \dots, L_n)$ is not $Pol(\mathcal{D})$-separable.*

Proof. We prove the contrapositive. Assume that $(\{\varepsilon\}, L_1, \dots, L_n)$ is $Pol(\mathcal{D})$-separable: we get $K \in Pol(\mathcal{D})$ such that $\varepsilon \in K$ and $(L_1, \dots, L_n) \cap K$ is $Pol(\mathcal{D})$-separable. By definition, K is a finite union of marked product of languages in $\mathcal{D}$. Hence, since $\varepsilon \in K$, there exists a marked product involving a single language $H \in \mathcal{D}$ such that $\varepsilon \in H$ in the union defining K. In particular, $H \subseteq K$ and Lemma 13 implies that $(L_1, \dots, L_n) \cap H$ is $Pol(\mathcal{D})$-separable. Since $H \in \mathcal{D}$ and $\varepsilon \in H$, it follows that $(L_1, \dots, L_n)$ is $Pol(\mathcal{D})$-separable under $\mathcal{D}$-control. ◀

▶ **Lemma 17.** *Let $\mathcal{D}$ be a prevariety and $w \in A^+$. If $(L_1, \dots, L_n)$ is not $Pol(\mathcal{D}^+)$-separable under $\mathcal{D}$-control, then $(w^+, w^+L_1w^+, \dots, w^+L_nw^+)$ is not $Pol(\mathcal{D}^+)$-separable.*

Proof. We prove the contrapositive. Assuming that $(w^+, w^+L_1w^+, \dots, w^+L_nw^+)$ is $Pol(\mathcal{D}^+)$-separable, we show that $(L_1, \dots, L_n)$ is $Pol(\mathcal{D}^+)$-separable under $\mathcal{D}$-control. There exists $K \in Pol(\mathcal{D}^+)$ such that $w^+ \subseteq K$, and $(w^+L_1w^+, \dots, w^+L_nw^+) \cap K$ is $Pol(\mathcal{D}^+)$-separable. By definition, K is a finite union of marked products $K_0a_1K_1 \cdots a_mK_m$ with $a_1, \dots, a_m \in A$ and $K_0, \dots, K_m \in \mathcal{D}^+$. Let $k \in \mathbb{N}$ such that $m \leq k$ for every product $K_0a_1K_1 \cdots a_mK_m$ in this union. Since $w^+ \subseteq K$, we have $w^{2(k+1)} \in K$. This yields a marked product $K_0a_1K_1 \cdots a_mK_m$ such that $w^{2(k+1)} \in K_0a_1K_1 \cdots a_mK_m \subseteq K$, $m \leq k$ and $K_0, \dots, K_m \in \mathcal{D}^+$. Therefore, we get $u_i \in K_i$ for each $i \leq m$ such that $w^{2(k+1)} = u_0a_1u_1 \cdots a_mu_m$. Moreover, since $m \leq k$, there exists $i \leq m$ such that ww is an infix of u_i. Thus, we get $x, y \in A^*$ and $\ell_1, \ell_2 \in \mathbb{N}$ such that $u_i = xwwy$, $u_0a_1u_1 \cdots a_ix = w^{\ell_1}$, $ya_{i+1}u_{i+1} \cdots a_mu_m = w^{\ell_2}$ and $\ell_1 + 2 + \ell_2 = 2(k+1)$

By definition $K_i \in \mathcal{D}^+$ which yields $H \in \mathcal{D}$ such that either $K_i = H \cup \{\varepsilon\}$ or $K_i = H \cap A^+$. Hence, since $u_i \in K_i$ and $u_i \in A^+$ (recall that $w \in A^+$), we have $xwwy = u_i \in H$. Let $H' = (xw)^{-1}H(wy)^{-1}$. By closure under quotients, we have $H' \in \mathcal{D}$ and it is clear that $\varepsilon \in H'$ since $xwwy \in H$. Hence, it remains to prove that $(L_1, \dots, L_n) \cap H'$ is $Pol(\mathcal{D}^+)$-separable. This will imply as desired that $(L_1, \dots, L_n)$ is $Pol(\mathcal{D}^+)$-separable under $\mathcal{D}$-control.

We know that $(w^+L_1w^+, \dots, w^+L_nw^+) \cap K$ is $Pol(\mathcal{D}^+)$-separable. One may verify from the definitions that $w^{\ell_1+1}(L_j \cap H')w^{\ell_2+1} \subseteq w^+L_jw^+ \cap K$ for all $j \leq n$. Thus, Lemma 13 implies that $w^{\ell_1+1}(L_1 \cap H')w^{\ell_2+1}, \dots, w^{\ell_1+1}(L_n \cap H')w^{\ell_2+1})$ is $Pol(\mathcal{D}^+)$-separable. Finally, since $(w^{\ell_1+1}, \dots, w^{\ell_1+1})$ and $(w^{\ell_2+1}, \dots, w^{\ell_2+1})$ are not $Pol(\mathcal{D}^+)$-separable, it follows from Lemma 14 that $((L_1 \cap H'), \dots, (L_n \cap H'))$ is $Pol(\mathcal{D}^+)$-separable as desired. ◀

4 Separation Algorithms for $BPol(\mathcal{G})$ and $BPol(\mathcal{G}^+)$

For a group prevariety $\mathcal{G}$, we now consider $BPol(\mathcal{G})$- and $BPol(\mathcal{G}^+)$-separation. We rely on the notions of Section 3: given an arbitrary NFA $\mathcal{A} = (Q, \delta)$, we present a generic characterization of the inseparable $BPol(\mathcal{G})$- and $BPol(\mathcal{G}^+)$-quadruples under $\mathcal{G}$ control associated to $\mathcal{A}$, *i.e.*, of the subsets $\mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}]$ and $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$ of Q^4. Thanks to Proposition 12, this also yields characterizations of $\mathcal{I}_{BPol(\mathcal{G})}[\mathcal{A}]$ and of $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{A}]$, which in turn, in view of Proposition 9, yield reductions from both $BPol(\mathcal{G})$- and $BPol(\mathcal{G}^+)$-separation to $\mathcal{G}$-separation. These polynomial time reductions are therefore *effective* when $\mathcal{G}$-separation is decidable.

4.1 Statements

Let $\mathcal{G}$ be a group prevariety and let $\mathcal{A} = (Q, \delta)$ be an NFA. We present characterizations of $\mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}]$ and $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$. They follow the same pattern, but each of them depends on a specific function from 2^{Q^4} to 2^{Q^4}, which we first describe.

Characterization of $\mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}]$. We use a function $\tau_{\mathcal{A},\mathcal{G}} : 2^{Q^4} \to 2^{Q^4}$. For $S \subseteq Q^4$, we define the set $\tau_{\mathcal{A},\mathcal{G}}(S) \subseteq Q^4$. The definition is based on an auxiliary NFA $\mathcal{B}_S = (Q^3, \gamma_S)$ *with ε-transitions*, which depends on S. Its states are triples in Q^3. The set $\gamma_S \subseteq Q^3 \times (A \cup \{\varepsilon\}) \times Q^3$ includes two kinds of transitions. First, given $a \in A$ and $s_1, s_2, s_3, t_1, t_2, t_3 \in Q$, we let $\big((s_1, s_2, s_3), a, (t_1, t_2, t_3)\big) \in \gamma_S$ if and only if $(s_1, a, t_1) \in \delta$, $(s_2, a, t_2) \in \delta$ and $(s_3, a, t_3) \in \delta$. Second, for every state $q_1 \in Q$ and every $(q_2, r_2, q_3, r_3) \in S$, we add the following ε-transition: $((q_1, q_2, q_3), \varepsilon, (q_1, r_2, r_3)) \in \gamma_S$. We represent this construction process graphically in Figure 1.

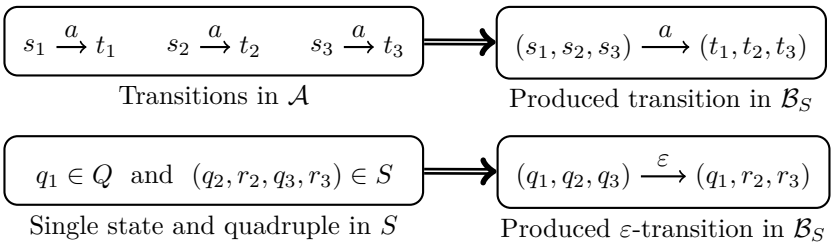

Figure 1 Construction of the transitions in the auxiliary automaton $\mathcal{B}_S$

▶ Remark 18. The NFA $\mathcal{B}_S$ and its counterpart $\mathcal{B}_S^+$ (which we define below as a means to handle $BPol(\mathcal{G}^+)$) are the *only* NFAs with ε-transitions considered in the paper. In particular, the original input NFA $\mathcal{A}$ is assumed to be *without* ε-transitions.

We are ready to define $\tau_{\mathcal{A},\mathcal{G}}(S) \subseteq Q^4$. For every $(q, r, s, t) \in Q^4$, we let $(q, r, s, t) \in \tau_{\mathcal{A},\mathcal{G}}(S)$ if and only if the two following conditions hold:

$$\begin{array}{l} \{\varepsilon\} \text{ is } \textit{not } \mathcal{G}\text{-separable from } L_{\mathcal{B}_S}((s, q, s), (t, r, t)), \text{ and} \\ \{\varepsilon\} \text{ is } \textit{not } \mathcal{G}\text{-separable from } L_{\mathcal{B}_S}((q, s, q), (r, t, r)). \end{array} \tag{1}$$

A set $S \subseteq Q^4$ is $(BPol, *)$*-sound* for $\mathcal{G}$ and $\mathcal{A}$ if it is a fixpoint for $\tau_{\mathcal{A},\mathcal{G}}$, *i.e.* $\tau_{\mathcal{A},\mathcal{G}}(S) = S$. We have the following simple lemma which can be verified from the definition. It states that $\tau_{\mathcal{A},\mathcal{G}} : 2^{Q^4} \to 2^{Q^4}$ is *increasing* (for inclusion). In particular, this implies that it has a *greatest fixpoint*, *i.e.*, there is a *greatest* $(BPol, *)$*-sound set.*

▶ **Lemma 19.** *Let $\mathcal{G}$ be a group prevariety and let $\mathcal{A} = (Q, \delta)$ be an NFA. For every $S, S' \subseteq Q^4$, we have $S \subseteq S' \Rightarrow \tau_{\mathcal{A},\mathcal{G}}(S) \subseteq \tau_{\mathcal{A},\mathcal{G}}(S')$.*

We may now state the first key theorem of the paper. It applies to $BPol(\mathcal{G})$-separation.

▶ **Theorem 20.** *Let $\mathcal{G}$ be a group prevariety and $\mathcal{A} = (Q, \delta)$ an NFA. Then, $\mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}]$ is the greatest $(BPol, *)$-sound subset of Q^4 for $\mathcal{G}$ and $\mathcal{A}$.*

Characterization of $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$. The characterization of $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$ is analogous. Roughly, the only difference is that we modify the definition of the auxiliary automaton $\mathcal{B}_S$. Let $\mathcal{G}$ be a group prevariety and $\mathcal{A} = (Q, \delta)$ be an NFA. We define a new function $\tau^+_{\mathcal{A},\mathcal{G}} : 2^{Q^4} \to 2^{Q^4}$. For $S \subseteq Q^4$, we define $\tau^+_{\mathcal{A},\mathcal{G}}(S) \subseteq Q^4$ using another auxiliary NFA $\mathcal{B}_S^+ = (Q^3, \gamma_S^+)$ *with ε-transitions*. Its states are triples in Q^3 and $\gamma_S^+ \subseteq Q^3 \times (A \cup$

$\{\varepsilon\}) \times Q^3$ contains two kinds of transitions. First, for $a \in A$ and $s_1, s_2, s_3, t_1, t_2, t_3 \in Q$, we let $\big((s_1, s_2, s_3), a, (t_1, t_2, t_3)\big) \in \gamma_S^+$ if and only if $(s_1, a, t_1) \in \delta$, $(s_2, a, t_2) \in \delta$ and $(s_3, a, t_3) \in \delta$. Second, for all $q_1 \in Q$ and all $(q_2, r_2, q_3, r_3) \in S$, if $A^+ \cap L_{\mathcal{A}}(q_1, q_1) \cap L_{\mathcal{A}}(q_2, q_2) \cap L_{\mathcal{A}}(q_3, q_3) \cap L_{\mathcal{A}}(r_2, r_2) \cap L_{\mathcal{A}}(r_3, r_3) \neq \emptyset$, then we add the following ε-transition: $((q_1, q_2, q_3), \varepsilon, (q_1, r_2, r_3)) \in \gamma_S^+$. We represent this construction in Figure 2.

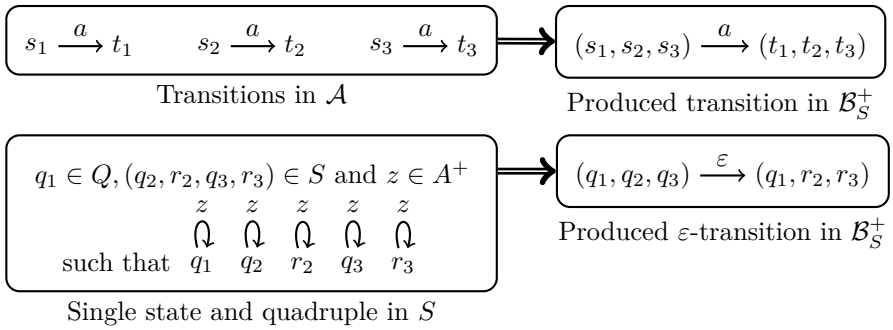

Figure 2 Construction of the transitions in the auxiliary automaton $\mathcal{B}_S^+$

We are ready to define $\tau^+_{\mathcal{A},\mathcal{G}}(S) \subseteq Q^4$. For every $(q, r, s, t) \in Q^4$, we let $(q, r, s, t) \in \tau^+_{\mathcal{A},\mathcal{G}}(S)$ if and only if the two following conditions hold:

$$\begin{array}{l} \{\varepsilon\} \text{ is } \textit{not } \mathcal{G}\text{-separable from } L_{\mathcal{B}_S^+}((s, q, s), (t, r, t)), \text{ and} \\ \{\varepsilon\} \text{ is } \textit{not } \mathcal{G}\text{-separable from } L_{\mathcal{B}_S^+}((q, s, q), (r, t, r)). \end{array} \tag{2}$$

A set $S \subseteq Q^4$ is $(BPol, +)$-*sound* for $\mathcal{G}$ and $\mathcal{A}$ if it is a fixpoint for $\tau^+_{\mathcal{A},\mathcal{G}}$, *i.e.* $\tau^+_{\mathcal{A},\mathcal{G}}(S) = S$. The following monotonicity lemma implies that there is a *greatest* $(BPol, +)$-*sound set*.

▶ **Lemma 21.** *Let $\mathcal{G}$ be a group prevariety and $\mathcal{A} = (Q, \delta)$ an NFA. For every $S, S' \subseteq Q^4$, we have $S \subseteq S' \Rightarrow \tau^+_{\mathcal{A},\mathcal{G}}(S) \subseteq \tau^+_{\mathcal{A},\mathcal{G}}(S')$.*

We may now state our second key theorem. It applies to $BPol(\mathcal{G}^+)$-separation.

▶ **Theorem 22.** *Let $\mathcal{G}$ be a group prevariety and $\mathcal{A} = (Q, \delta)$ an NFA. Then, $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$ is the greatest $(BPol, +)$-sound subset of Q^4 for $\mathcal{G}$ and $\mathcal{A}$.*

Let us discuss the consequences of Theorems 20 and 22. Since $\mathcal{B}_S$ and $\mathcal{B}_S^+$ can be computed from $\mathcal{A}$ and S, one can compute $\tau_{\mathcal{A},\mathcal{G}}(S)$ and $\tau^+_{\mathcal{A},\mathcal{G}}(S)$ from S provided that $\mathcal{G}$-separation is decidable. Hence, if $\mathcal{G}$-separation is decidable, Theorem 20 (resp. Theorem 22) yields a *greatest* fixpoint procedure for computing $\mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}]$ (resp. $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$). Indeed, consider the sequence of subsets defined by $S_0 = Q^4$, and $S_n = \tau_{\mathcal{A},\mathcal{G}}(S_{n-1})$ for $n \geq 1$. By definition, computing S_n from S_{n-1} boils down to deciding $\mathcal{G}$-separation. Since $\tau_{\mathcal{A},\mathcal{G}}$ is increasing by Lemma 19, we get a decreasing sequence $Q^4 = S_0 \supseteq S_1 \supseteq S_2 \cdots$. Moreover, since Q^4 is finite, this sequence stabilizes at some point: there exists $n \in \mathbb{N}$ such that $S_n = S_j$ for all $j \geq n$. One may verify that S_n is the greatest $(BPol, *)$-sound subset of Q^4. By Theorem 20, it follows that $S_n = \mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}]$. Likewise, the sequence T_n defined by $T_0 = Q^4$ and $T_n = \tau^+_{\mathcal{A},\mathcal{G}}(T_{n-1})$ is computable when $\mathcal{G}$-separation is decidable, and, since it is decreasing, it stabilizes. By Theorem 22, its stabilization value is $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$.

By Proposition 12, $\mathcal{I}_{BPol(\mathcal{G})}[\mathcal{A}]$ (resp. $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{A}]$) can be computed from $\mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}]$ (resp. $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$) via a *least* fixpoint procedure. Altogether, by Proposition 9, we get reductions from $BPol(\mathcal{G})$- and $BPol(\mathcal{G}^+)$-separation to $\mathcal{G}$-separation. One may verify that

these are polynomial time reductions (we mean "reduction" in the Turing sense: $BPol(\mathcal{G})$- and $BPol(\mathcal{G}^+)$-separation can be decided in polynomial time using an oracle for $\mathcal{G}$-separation).

Now, it is known that separation can be decided in polynomial time for the classes ST, MOD and GR (this is trivial for ST, see [26] for MOD and GR). Hence, we obtain from Theorem 20 that separation is decidable in polynomial time for $BPol(\mathrm{ST})$ (*i.e.*, $\mathcal{B}\Sigma_1(<)$), $BPol(\mathrm{MOD})$ (*i.e.*, $\mathcal{B}\Sigma_1(<, MOD)$) and $BPol(\mathrm{GR})$. This was well-know for $BPol(\mathrm{ST})$ (the class of piecewise testable languages, see [6, 17]). For the other two, decidability was known [36, 21] but not the polynomial time upper bound. Using Theorem 22, we also obtain that separation is decidable in polynomial time for $BPol(\mathrm{ST}^+)$ (*i.e.*, the languages of dot-depth one or equivalently $\mathcal{B}\Sigma_1(<, +1)$), $BPol(\mathrm{MOD}^+)$ (*i.e.*, $\mathcal{B}\Sigma_1(<, +1, MOD)$) and $BPol(\mathrm{GR}^+)$. Decidability was already known for $BPol(\mathrm{ST}^+)$ and $BPol(\mathrm{MOD}^+)$: the results can be obtained indirectly by reduction to $BPol(\mathrm{ST})$-separation using transfer theorems [22, 16]. Yet, the polynomial time upper bounds are new as the transfer theorems have a built-in exponential blow-up. Moreover, decidability of separation is a new result for $BPol(\mathrm{GR}^+)$.

Finally, the statement applies to $BPol(\mathrm{AMT})$ and $BPol(\mathrm{AMT}^+)$ (*i.e.*, $\mathcal{B}\Sigma_1(<, AMOD)$ and $\mathcal{B}\Sigma_1(<, +1, AMOD)$). This is a new result for $BPol(\mathrm{AMT}^+)$. Yet, since AMT-separation is co-NP-complete when the alphabet is part of the input [26] (the problem being in P for a fixed alphabet), the complexity analysis is not entirely immediate. However, one may verify that the procedures yield co-NP algorithms for both $BPol(\mathrm{AMT})$- and $BPol(\mathrm{AMT}^+)$-separation. We summarize the upper bounds in Figure 3.

Input class $\mathcal{G}$	ST	MOD	AMT	GR
$BPol(\mathcal{G})$- and $BPol(\mathcal{G}^+)$-separation	P	P	co-NP	P

Figure 3 Complexity of separation (for input languages represented by NFAs).

4.2 Proof of Theorem 20

We now concentrate on the proof of Theorem 20. The key ingredients in this argument are Proposition 6 and Lemma 16. The proof of Theorem 22 is available in the appendix. It is based on similar ideas. Roughly, we replace Proposition 6 and Lemma 16 (which are tailored to classes $BPol(\mathcal{G})$) by their counterparts for $BPol(\mathcal{G}^+)$: Proposition 7 and Lemma 17. However, note that proving Theorem 22 is technically more involved as manipulating the automaton $\mathcal{B}_S^+$ in the definition of $\tau^+_{\mathcal{A},\mathcal{G}}$ requires more work.

We fix a group prevariety $\mathcal{G}$ and an NFA $\mathcal{A} = (Q, \delta)$. Let $S \subseteq Q^4$ be the greatest $(BPol, *)$-sound subset for $\mathcal{G}$ and $\mathcal{A}$. We prove that $S = \mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}]$.

First part: $S \subseteq \mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}]$**.** We use *tuple separation* and Lemma 16. Let us start with some terminology. For every $n \geq 1$ and $(q_1, r_1, q_2, r_2) \in Q^4$, we associate an n-tuple of languages, written $T_n(q_1, r_1, q_2, r_2)$. We use induction on n and tuple concatenation to present the definition. If $n = 1$ then, $T_1(q_1, r_1, q_2, r_2) = \big(L_{\mathcal{A}}(q_2, r_2)\big)$. If $n > 1$, then,

$$T_n(q_1, r_1, q_2, r_2) = \begin{cases} (L_{\mathcal{A}}(q_2, r_2)) \cdot T_{n-1}(q_1, r_1, q_2, r_2) & \text{if } n \text{ is odd} \\ (L_{\mathcal{A}}(q_1, r_1)) \cdot T_{n-1}(q_1, r_1, q_2, r_2) & \text{if } n \text{ is even.} \end{cases}$$

For example, we have $T_3(q_1, r_1, q_2, r_2) = (L_{\mathcal{A}}(q_2, r_2), L_{\mathcal{A}}(q_1, r_1), L_{\mathcal{A}}(q_2, r_2))$.

▶ **Proposition 23.** *For every $n \geq 1$ and $(q_1, r_1, q_2, r_2) \in S$, the n-tuple $T_n(q_1, r_1, q_2, r_2)$ is not $Pol(\mathcal{G})$-separable under $\mathcal{G}$-control.*

By definition, Proposition 23 implies that for all $p \geq 1$ and $(q_1, r_1, q_2, r_2) \in S$, the $2p$-tuple $(L_{\mathcal{A}}(q_1, r_1), L_{\mathcal{A}}(q_2, r_2))^p$ is not $Pol(\mathcal{G})$-separable under $\mathcal{G}$-control. By Corollary 15, it follows that $L_{\mathcal{A}}(q_1, r_1)$ is not $BPol(\mathcal{G})$-separable from $L_{\mathcal{A}}(q_2, r_2)$ under $\mathcal{G}$-control, *i.e.*, that $(q_1, r_1, q_2, r_2) \in \mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}]$. We get $S \subseteq \mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}]$ as desired.

We prove Proposition 23 by induction on n. We fix $n \geq 1$ for the proof. In order to exploit the hypothesis that S is $(BPol, *)$-sound, we need a property of the NFA $\mathcal{B}_S = (Q^3, \gamma_S)$ used to define $\tau_{\mathcal{A},\mathcal{G}}$. When $n \geq 2$, this is where we use induction on n and Lemma 16.

▶ **Lemma 24.** *Let $(s_1, s_2, s_3), (t_1, t_2, t_3) \in Q^3$ and $w \in L_{\mathcal{B}_S}((s_1, s_2, s_3), (t_1, t_2, t_3))$. Then, $w \in L_{\mathcal{A}}(s_1, t_1)$ and, if $n \geq 2$, the n-tuple $(\{w\}) \cdot T_{n-1}(s_2, t_2, s_3, t_3)$ is not $Pol(\mathcal{G})$-separable.*

Proof. Since $w \in L_{\mathcal{B}_S}((s_1, s_2, s_3), (t_1, t_2, t_3))$, there exists a run labeled by w from (s_1, s_2, s_3) to (t_1, t_2, t_3) in $\mathcal{B}_S$. We use a sub-induction on the number of transitions involved in that run. First, assume that no transitions are used: we have $w = \varepsilon$ and $(s_1, s_2, s_3) = (t_1, t_2, t_3)$. Clearly, $\varepsilon \in L_{\mathcal{A}}(s_1, s_1)$ and, if $n \geq 2$, the n-tuple $(\{\varepsilon\}) \cdot T_{n-1}(s_2, s_2, s_3, s_3)$ is not $Pol(\mathcal{G})$-separable by Lemma 13 since $\varepsilon \in L_{\mathcal{A}}(s_2, s_2) \cap L_{\mathcal{A}}(s_3, s_3)$. We now assume that at least one transition is used and consider the last one: we have $(q_1, q_2, q_3) \in Q^3$, $w' \in A^*$ and $x \in A \cup \{\varepsilon\}$ such that $w = w'x$, $w' \in L_{\mathcal{B}_S}((s_1, s_2, s_3), (q_1, q_2, q_3))$ and $((q_1, q_2, q_3), x, (t_1, t_2, t_3)) \in \gamma_S$. By induction, we have $w' \in L_{\mathcal{A}}(s_1, q_1)$ and, if $n \geq 2$, the n-tuple $(\{w'\}) \cdot T_{n-1}(s_2, q_2, s_3, q_3)$ is not $Pol(\mathcal{G})$-separable. We prove that $x \in L_{\mathcal{A}}(q_1, t_1)$ and, if $n \geq 2$, the n-tuple $(\{x\}) \cdot T_{n-1}(q_2, t_2, q_3, t_3)$ is not $Pol(\mathcal{G})$-separable. It will then be immediate that $w = w'x \in L_{\mathcal{A}}(s_1, t_1)$ and, if $n \geq 2$, Lemma 14 implies that $(\{w\}) \cdot T_{n-1}(s_2, t_2, s_3, t_3)$ is not $Pol(\mathcal{G})$-separable.

We consider two cases depending on whether $x \in A$ or $x = \varepsilon$. First, if $x = a \in A$, then $(q_i, a, t_i) \in \delta$ for $i = \{1, 2, 3\}$. Clearly, this implies that $a \in L_{\mathcal{A}}(q_1, t_1)$ and, if $n \geq 2$, then $(\{a\}) \cdot T_{n-1}(q_2, t_2, q_3, t_3)$ is not $Pol(\mathcal{G})$-separable by Lemma 13 since $a \in L_{\mathcal{A}}(q_2, t_2) \cap L_{\mathcal{A}}(q_3, t_3)$. Assume now that $x = \varepsilon$: we are dealing with an ε-transition. By definition of γ_S, we have $q_1 = t_1$ and $(q_2, t_2, q_3, t_3) \in S$. The former yields $\varepsilon \in L_{\mathcal{A}}(q_1, t_1)$. Moreover, if $n \geq 2$, since $(q_2, t_2, q_3, t_3) \in S$, it follows from induction on n in Proposition 23 that the $(n-1)$-tuple $T_{n-1}(q_2, t_2, q_3, t_3)$ is not $Pol(\mathcal{G})$-separable under $\mathcal{G}$-control. Combined with Lemma 16, this yields that $(\{\varepsilon\}) \cdot T_{n-1}(q_2, t_2, q_3, t_3)$ is not $Pol(\mathcal{G})$-separable, as desired. ◀

We may now complete the proof of Proposition 23. By symmetry, we only treat the case when n is odd and leave the case when it is even to the reader. Let $(q_1, r_1, q_2, r_2) \in S$, we have to prove that $T_n(q_1, r_1, q_2, r_2)$ is not $Pol(\mathcal{G})$-separable under $\mathcal{G}$-control. Hence, we fix $H \in \mathcal{G}$ such that $\varepsilon \in H$ and prove $H \cap T_n(q_1, r_1, q_2, r_2)$ is not $Pol(\mathcal{G})$-separable. Since S is $(BPol, *)$-sound, we have $\tau_{\mathcal{A},\mathcal{G}}(S) = S$, which implies that $(q_1, r_1, q_2, r_2) \in \tau_{\mathcal{A},\mathcal{G}}(S)$. Hence, it follows from (1) that $\{\varepsilon\}$ is not $\mathcal{G}$-separable from $L_{\mathcal{B}_S}((q_2, q_1, q_2), (r_2, r_1, r_2))$. Since $H \in \mathcal{G}$ and $\varepsilon \in H$, we get a word $w \in H \cap L_{\mathcal{B}_S}((q_2, q_1, q_2), (r_2, r_1, r_2))$. By Lemma 24, we have $w \in H \cap L_{\mathcal{A}}(q_2, r_2)$. This completes the proof when $n = 1$. Indeed, in that case we have $T_1(q_1, r_1, q_2, r_2) = (L_{\mathcal{A}}(q_2, r_2))$ and since $H \cap L_{\mathcal{A}}(q_2, r_2) \neq \emptyset$, it follows that $H \cap T_1(q_1, r_1, q_2, r_2)$ is not $Pol(\mathcal{G})$-separable, as desired. If $n \geq 2$, then Lemma 24 also implies that $(\{w\}) \cdot T_{n-1}(q_1, r_1, q_2, r_2)$ is not $Pol(\mathcal{G})$-separable. Since $w \in H \cap L_{\mathcal{A}}(q_2, r_2)$, Lemma 13 yields that $(H \cap L_{\mathcal{A}}(q_2, r_2)) \cdot T_{n-1}(q_1, r_1, q_2, r_2)$ is not $Pol(\mathcal{G})$-separable. Thus, since $H \in \mathcal{G} \subseteq Pol(\mathcal{G})$, one may verify that the n-tuple $(H \cap L_{\mathcal{A}}(q_2, r_2)) \cdot (H \cap T_{n-1}(q_1, r_1, q_2, r_2))$ is not $Pol(\mathcal{G})$-separable. By definition, this exactly says that $H \cap T_n(q_1, r_1, q_2, r_2)$ is not $Pol(\mathcal{G})$-separable, completing the proof.

Second part: $\mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}] \subseteq S$. In the sequel, we say that an arbitrary set $R \subseteq Q^4$ is *good* if there exists $L \in \mathcal{G}$ such $\varepsilon \in L$ and a $BPol(\mathcal{G})$-cover $\mathbf{K}$ of L which is separating for R.

▶ **Proposition 25.** *Let $R \subseteq Q^4$. If R is good, then $\tau_{\mathcal{A},\mathcal{G}}(R)$ is good as well.*

We use Proposition 25 to complete the proof. Let $S_0 = Q^4$ and $S_i = \tau_{\mathcal{A},\mathcal{G}}(S_{i-1})$ for $i \geq 1$. By Lemma 19, we have $S_0 \supseteq S_1 \subseteq S_2 \supseteq \cdots$ and there is $n \in \mathbb{N}$ such that S_n is the greatest $(BPol, *)$-sound subset for $\mathcal{G}$ and $\mathcal{A}$, *i.e.*, such that $S_n = S$. Since S_0 is good ($\{A^*\}$ is a $BPol(\mathcal{G})$-cover of $A^* \in \mathcal{G}$ which is separating for $S_0 = Q^4$), Proposition 25 implies that S_i is good for all $i \in \mathbb{N}$. Thus, $S = S_n$ is good. We get $L \in \mathcal{G}$ such that $\varepsilon \in L$ and a $BPol(\mathcal{G})$-cover $\mathbf{K}$ of L which is separating for S. Lemma 11 then yields $\mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}] \subseteq S$ as desired.

▶ Remark 26. The proof of Proposition 25 actually provides a construction for building $L \in \mathcal{G}$ such that $\varepsilon \in L$ and a $BPol(\mathcal{G})$-cover $\mathbf{K}$ of L which is separating for S (yet, this involves building separators in $\mathcal{G}$, see Lemma 27). As we have now established that $S = \mathcal{I}_{BPol(\mathcal{G})}[\mathcal{G}, \mathcal{A}]$, one may then follow the proof of Proposition 12 to build a $BPol(\mathcal{G})$-cover $\mathbf{H}$ of A^* which is separating for $\mathcal{I}_{BPol(\mathcal{G})}[\mathcal{A}]$. Finally, $\mathbf{H}$ encodes separators for all pairs of languages recognized by $\mathcal{A}$ which are $BPol(\mathcal{G})$-separable (roughly, this is the proof of Lemma 10). Altogether, we get a way to build separators in $BPol(\mathcal{G})$, when they exist.

We now prove Proposition 25. Let $R \subseteq Q^4$ be good. We have to build $L \in \mathcal{G}$ with $\varepsilon \in L$ and a $BPol(\mathcal{G})$-cover $\mathbf{K}$ of L which is separating for $\tau_{\mathcal{A},\mathcal{G}}(R)$ (which will prove that $\tau_{\mathcal{A},\mathcal{G}}(R)$ is good as well). We first build L (this part is independent from our hypothesis on R).

▶ **Lemma 27.** *There exists $L \in \mathcal{G}$ such that $\varepsilon \in L$ and for every $(q, r, s, t) \in Q^4$, if $L_{\mathcal{B}_R}((q, s, q), (r, t, r)) \cap L \neq \emptyset$ and $L_{\mathcal{B}_R}((s, q, s), (t, r, t)) \cap L \neq \emptyset$, then $(q, r, s, t) \in \tau_{\mathcal{A},\mathcal{G}}(R)$.*

Proof. Let $\mathbf{H}$ be the *finite* set of all languages recognized by $\mathcal{B}_R$ such that $\{\varepsilon\}$ is $\mathcal{G}$-separable from H. For every $H \in \mathbf{H}$, there exists $L_H \in \mathcal{G}$ such that $\varepsilon \in L_H$ and $L_H \cap H = \emptyset$. We define $L = \bigcap_{H \in \mathbf{H}} L_H \in \mathcal{G}$. It is clear that $\varepsilon \in L$. Moreover, given $(q, r, s, t) \in Q^4$, if $L_{\mathcal{B}_R}((q, s, q), (r, t, r)) \cap L \neq \emptyset$ and $L_{\mathcal{B}_R}((s, q, s), (t, r, t)) \cap L \neq \emptyset$, it follows from the definition of L that $\{\varepsilon\}$ is not $\mathcal{G}$-separable from both $L_{\mathcal{B}_R}((q, s, q), (r, t, r))$ and $L_{\mathcal{B}_R}((s, q, s), (t, r, t))$. It follows from (1) in the definition of $\tau_{\mathcal{A},\mathcal{G}}$ that $(q, r, s, t) \in \tau_{\mathcal{A},\mathcal{G}}(R)$. ◀

We fix $L \in \mathcal{G}$ as described in Lemma 27 for the remainder of the proof. We now build the $BPol(\mathcal{G})$-cover $\mathbf{K}$ of L using the hypothesis that R is good and Proposition 6.

▶ **Lemma 28.** *For all $(q, r) \in Q^2$, there is $H_{q,r} \in BPol(\mathcal{G})$ such that $L_{\mathcal{A}}(q, r) \cap L \subseteq H_{q,r}$ and for all pairs $(s, t) \in Q^2$, if $L_{\mathcal{A}}(s, t) \cap H_{q,r} \neq \emptyset$ then $L_{\mathcal{B}_R}((q, s, q), (r, t, r)) \cap L \neq \emptyset$.*

Proof. Since R is good, there are $U \in \mathcal{G}$ such that $\varepsilon \in U$ and a $BPol(\mathcal{G})$-cover $\mathbf{V}$ of U which is separating for R. We use them to build $H_{q,r}$. Since U is a group language and $\varepsilon \in U$, Proposition 6 yields a cover $\mathbf{P}$ of $L_{\mathcal{A}}(q, r) \cap L$ such that every $P \in \mathbf{P}$ is of the form $P = \uparrow_U w_P$ where $w_P \in L_{\mathcal{A}}(q, r) \cap L$. For every $P \in \mathbf{P}$, we build a $BPol(\mathcal{G})$-cover $\mathbf{K}_P$ of P. Let $a_1, \ldots, a_n \in A$ be the letters such that $w_P = a_1 \cdots a_n$. We have $P = Ua_1U \cdots a_nU$. Since $U \in \mathcal{G} \subseteq Pol(\mathcal{G})$ and $\mathbf{V}$ is a $BPol(\mathcal{G})$-cover of U, Proposition 5 yields a $BPol(\mathcal{G})$-cover $\mathbf{K}_P$ of P such that for every $K \in \mathbf{K}_P$, there exist $V_0, \ldots, V_n \in \mathbf{V}$ satisfying $K \subseteq V_0a_1V_1 \cdots a_nV_n$. We define $H_{q,r}$ as the union of all languages K such that $K \in \mathbf{K}_P$ for some $P \in \mathbf{P}$ and $L_{\mathcal{A}}(q, r) \cap K \neq \emptyset$. Clearly, $H_{q,r} \in BPol(\mathcal{G})$. Moreover, since $\mathbf{P}$ is a cover of $L_{\mathcal{A}}(q, r) \cap L$, and $\mathbf{K}_P$ is a cover of P for each $P \in \mathbf{P}$, it is clear that $L_{\mathcal{A}}(q, r) \cap L \subseteq H_{q,r}$. We now fix $(s, t) \in Q^2$ such that $L_{\mathcal{A}}(s, t) \cap H_{q,r} \neq \emptyset$ and show that $L_{\mathcal{B}_R}((q, s, q), (r, t, r)) \cap L \neq \emptyset$. By definition of $H_{q,r}$, we get $P \in \mathbf{P}$ and $K \in \mathbf{K}_P$ such that $L_{\mathcal{A}}(q, r) \cap K \neq \emptyset$ and $L_{\mathcal{A}}(s, t) \cap K \neq \emptyset$. By definition, $P = \uparrow_U w_P$ with $w_P \in L_{\mathcal{A}}(q, r) \cap L$. Hence, it suffices to prove that $w_P \in L_{\mathcal{B}_R}((q, s, q), (r, t, r))$.

We fix $x \in L_{\mathcal{A}}(s, t) \cap K$ and $y \in L_{\mathcal{A}}(q, r) \cap K$. Recall that $w_P = a_1 \cdots a_n$ (if $n = 0$, then $w_P = \varepsilon$). Since $w_P \in L_{\mathcal{A}}(q, r)$, we may consider the corresponding run in $\mathcal{A}$: we get

$p_0, \dots, p_n \in Q$ such that $p_0 = q$, $p_n = r$ and $(p_{i-1}, a_i, p_i) \in \delta$ for $1 \leq i \leq n$. Moreover, since $K \in \mathbf{K}_P$ and $w_P = a_1 \cdots a_n$, we have $K \subseteq V_0 a_1 V_1 \cdots a_n V_n$ for $V_0, \dots, V_n \in \mathbf{V}$ (if $n = 0$, then $K \subseteq V_0$). Since $x, y \in K$, we get $x_i, y_i \in V_i$ for $0 \leq i \leq n$ such that $x = x_0 a_1 x_1 \cdots a_n x_n$ and $y = y_0 a_1 y_1 \cdots a_n y_n$. Since $x \in L_{\mathcal{A}}(s,t)$, we get $s_0, t_0, \dots, s_n, t_n \in Q$ such that $s_0 = s$, $t_n = t$, $x_i \in L_{\mathcal{A}}(s_i, t_i)$ for $0 \leq i \leq n$, and $(t_{i-1}, a_i, s_i) \in \delta$ for $1 \leq i \leq n$. Symmetrically, since $y \in L_{\mathcal{A}}(q,r)$, we get $q_0, r_0, \dots, q_n, r_n \in Q$ such that $q_0 = q$, $r_n = r$, $y_i \in L_{\mathcal{A}}(q_i, r_i)$ for $0 \leq i \leq n$, and $(r_{i-1}, a_i, q_i) \in \delta$ for $1 \leq i \leq n$. By definition of γ_R, it is immediate that $((p_{i-1}, t_{i-1}, r_{i-1}), a_i, (p_i, s_i, q_i)) \in \gamma_R$ for $1 \leq i \leq n$. Since $V_i \in \mathbf{V}$ and $\mathbf{V}$ is separating for R, the fact that $x_i, y_i \in V_i$ implies that $(s_i, t_i, q_i, r_i) \in R$ for $0 \leq i \leq n$. Hence, $((p_i, s_i, q_i), \varepsilon, (p_i, t_i, r_i)) \in \gamma_R$ by definition. Thus, we get a run labeled by w_P from (p_0, s_0, q_0) to (p_n, t_n, r_n) in $\mathcal{B}_R$, *i.e.*, $w_P \in L_{\mathcal{B}_R}((q,s,q),(r,t,r))$ as desired. ◀

We may now build $\mathbf{K}$. Let $\mathbf{H} = \{H_{q,r} \mid (q,r) \in Q^2\}$. Consider the following equivalence $\sim$ defined on L: given $u, v \in L$, we let $u \sim v$ if and only if $u \in H_{q,r} \Leftrightarrow v \in H_{q,r}$ for every $(q,r) \in Q^2$. We let $\mathbf{K}$ as the partition of L into $\sim$-classes. Clearly, each $K \in \mathbf{K}$ is a Boolean combination involving the languages in $\mathbf{H}$ (which belong to $BPol(\mathcal{G})$) and $L \in \mathcal{G}$. Hence, $\mathbf{K}$ is a $BPol(\mathcal{G})$-cover of L. We now prove that it is separating for $\tau_{\mathcal{A},\mathcal{G}}(R)$. Let $q, r, s, t \in Q$ and $K \in \mathbf{K}$ such that there are $u \in L_{\mathcal{A}}(q,r) \cap K$ and $v \in L_{\mathcal{A}}(s,t) \cap K$. We show that $(q,r,s,t) \in \tau_{\mathcal{A},\mathcal{G}}(R)$. By definition of $\mathbf{K}$, we have $u, v \in L$ and $u \sim v$. In particular, $u \in L_{\mathcal{A}}(q,r) \cap L$ which yields $u \in H_{q,r}$ by definition in Lemma 28. Together with $u \sim v$, this yields $v \in H_{q,r}$. Hence, $L_{\mathcal{A}}(s,t) \cap H_{q,r} \neq \emptyset$ and Lemma 28 yields $L_{\mathcal{B}_R}((q,s,q),(r,t,r)) \cap L \neq \emptyset$. One may now use a symmetrical argument to obtain $L_{\mathcal{B}_R}((s,q,s),(t,r,t)) \cap L \neq \emptyset$. By definition of L in Lemma 27, this yields $(q,r,s,t) \in \tau_{\mathcal{A},\mathcal{G}}(R)$, completing the proof.

5 Conclusion

In this paper, we proved that for every group prevariety $\mathcal{G}$, there exist generic polynomial time Turing reductions from $BPol(\mathcal{G})$- and $BPol(\mathcal{G}^+)$-separation to $\mathcal{G}$-separation, for input languages represented by NFAs. While a generic reduction from $BPol(\mathcal{G})$-separation to $\mathcal{G}$-separation was already developed in [21], it relied on an involved machinery, which required to dig into a more general problem than $BPol(\mathcal{G})$-separation, namely "$BPol(\mathcal{G})$-covering". In particular, the techniques from [21] do not provide any way to build separators in $BPol(\mathcal{G})$ (when they exist). They also yield poor upper complexity bounds. At last, the results of [21] do not apply to $BPol(\mathcal{G}^+)$. In this case, even the existence of a generic reduction is new. It would be interesting to unify ideas of the present paper with the techniques of [21], to lift them to the setting of $BPol(\mathcal{G})$- and $BPol(\mathcal{G}^+)$-covering. We leave this for further work.

Our results imply that separation is decidable in *polynomial time* for a number of standard classes: the piecewise testable languages (*i.e.*, $BPol(\mathrm{ST})$ or equivalently $\mathcal{B}\Sigma_1(<)$), the languages of dot-depth one (*i.e.*, $BPol(\mathrm{ST}^+)$ or equivalently $\mathcal{B}\Sigma_1(<,+1)$), the classes $BPol(\mathrm{MOD})$ and $BPol(\mathrm{MOD}^+)$ (*i.e.*, $\mathcal{B}\Sigma_1(<, MOD)$ and $\mathcal{B}\Sigma_1(<,+1, MOD)$) and the classes $BPol(\mathrm{GR})$ and $BPol(\mathrm{GR}^+)$. While this was well-known for the piecewise testable languages [17, 6], all other results are new—not only regarding the complexity, but even regarding the decidability. Actually, it is shown in [12] that $BPol(\mathrm{ST})$-separation is P-complete. It turns out that the reduction of [12], from the circuit value problem, adapts to prove the P-completeness of separation for all of the above classes (we leave the details for further work). Finally, our results also apply to the classes $BPol(\mathrm{AMT})$ and $BPol(\mathrm{AMT}^+)$ (*i.e.*, $\mathcal{B}\Sigma_1(<, AMOD)$ and $\mathcal{B}\Sigma_1(<,+1, AMOD)$): we obtain that separation is in co-NP. While this is currently unknown, we conjecture that this is a *tight* upper bound. Indeed, it is known that AMT-separation is co-NP-complete [26].

References

1. Jorge Almeida and Marc Zeitoun. The pseudovariety **J** is hyperdecidable. *RAIRO Theoretical Informatics and Applications*, 31(5):457–482, 1997.
2. Mustapha Arfi. Polynomial operations on rational languages. In *Proceedings of the 4th Annual Symposium on Theoretical Aspects of Computer Science*, STACS'87, pages 198–206, Berlin, Heidelberg, 1987. Springer-Verlag.
3. Janusz A. Brzozowski and Rina S. Cohen. Dot-depth of star-free events. *Journal of Computer and System Sciences*, 5(1):1–16, 1971.
4. Antonio Cano, Giovanna Guaiana, and Jean-Eric Pin. Regular languages and partial commutations. *Journal of Information and Computation*, 230:76–96, 2013.
5. Laura Chaubard, Jean Éric Pin, and Howard Straubing. First order formulas with modular predicates. In *Proceedings of the 21th IEEE Symposium on Logic in Computer Science (LICS'06)*, pages 211–220, 2006.
6. 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, Berlin, Heidelberg, 2013. Springer-Verlag.
7. Samuel Eilenberg. *Automata, Languages, and Machines*, volume B. Academic Press, Inc., Orlando, FL, USA, 1976.
8. Karsten Henckell, Stuart Margolis, Jean-Eric Pin, and John Rhodes. Ash's type II theorem, profinite topology and Malcev products. *International Journal of Algebra and Computation*, 1:411–436, 1991.
9. Robert Knast. A semigroup characterization of dot-depth one languages. *RAIRO - Theoretical Informatics and Applications*, 17(4):321–330, 1983.
10. Alexis Maciel, Pierre Péladeau, and Denis Thérien. Programs over semigroups of dot-depth one. *Theoretical Computer Science*, 245(1):135–148, 2000.
11. Stuart Margolis and Jean-Eric Pin. Product of Group Languages. In *FCT Conference*, volume Lecture Notes in Computer Science, pages 285–299. Springer-Verlag, 1985.
12. Tomás Masopust. Separability by piecewise testable languages is ptime-complete. *Theor. Comput. Sci.*, 711:109–114, 2018.
13. Jean-Eric Pin. Algebraic tools for the concatenation product. *Theoretical Computer Science*, 292:317–342, 2003.
14. Jean-Eric Pin. An explicit formula for the intersection of two polynomials of regular languages. In *DLT 2013*, volume 7907 of *Lect. Notes Comp. Sci.*, pages 31–45. Springer, 2013.
15. Jean-Eric Pin and Howard Straubing. Some results on $\mathcal{C}$-varieties. *RAIRO - Theoretical Informatics and Applications*, 39(1):239–262, 2005.
16. Thomas Place, Varun Ramanathan, and Pascal Weil. Covering and separation for logical fragments with modular predicates. *Logical Methods in Computer Science*, 15(2), 2019.
17. Thomas Place, Lorijn van Rooijen, and Marc Zeitoun. Separating regular languages by piecewise testable and unambiguous languages. In *Proceedings of the 38th International Symposium on Mathematical Foundations of Computer Science*, MFCS'13, pages 729–740, Berlin, Heidelberg, 2013. Springer-Verlag.
18. Thomas Place and Marc Zeitoun. Separation for dot-depth two. In *Proceedings of the 32th Annual ACM/IEEE Symposium on Logic in Computer Science, (LICS'17)*, pages 202–213. IEEE Computer Society, 2017.
19. Thomas Place and Marc Zeitoun. The covering problem. *Logical Methods in Computer Science*, 14(3), 2018.
20. Thomas Place and Marc Zeitoun. Generic results for concatenation hierarchies. *Theory of Computing Systems (ToCS)*, 63(4):849–901, 2019. Selected papers from CSR'17.
21. Thomas Place and Marc Zeitoun. Separation and covering for group based concatenation hierarchies. In *Proceedings of the 34th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS'19, pages 1–13, 2019.

719 **22** Thomas Place and Marc Zeitoun. Adding successor: A transfer theorem for separation and
720 covering. *ACM Transactions on Computational Logic*, 21(2):9:1–9:45, 2020.
721 **23** Thomas Place and Marc Zeitoun. Separation for dot-depth two. *Logical Methods in Computer*
722 *Science*, Volume 17, Issue 3, 2021.
723 **24** Thomas Place and Marc Zeitoun. Characterizing level one in group-based concatenation hier-
724 archies. In *Computer Science – Theory and Applications*, Cham, 2022. Springer International
725 Publishing.
726 **25** Thomas Place and Marc Zeitoun. A generic polynomial time approach to separation by first-
727 order logic without quantifier alternation, 2022. URL: `https://arxiv.org/abs/2210.00946`,
728 `doi:10.48550/ARXIV.2210.00946`.
729 **26** Thomas Place and Marc Zeitoun. Group separation strikes back. To appear, a preliminary
730 version is vailable at `https://www.labri.fr/perso/tplace/Files/groups.pdf`, 2022.
731 **27** Imre Simon. Piecewise testable events. In *Proceedings of the 2nd GI Conference on Automata*
732 *Theory and Formal Languages*, pages 214–222, Berlin, Heidelberg, 1975. Springer-Verlag.
733 **28** Benjamin Steinberg. Inevitable graphs and profinite topologies: Some solutions to algorithmic
734 problems in monoid and automata theory, stemming from group theory. *International Journal*
735 *of Algebra and Computation*, 11(1):25–72, 2001.
736 **29** Howard Straubing. A generalization of the schützenberger product of finite monoids. *Theoretical*
737 *Computer Science*, 13(2):137–150, 1981.
738 **30** Howard Straubing. Finite semigroup varieties of the form V $*$ D. *Journal of Pure and Applied*
739 *Algebra*, 36:53–94, 1985.
740 **31** Howard Straubing. On logical descriptions of regular languages. In *Proceedings of the 5th*
741 *Latin American Symposium on Theoretical Informatics*, LATIN'02, pages 528–538, Berlin,
742 Heidelberg, 2002. Springer-Verlag.
743 **32** Denis Thérien. Classification of finite monoids: The language approach. *Theoretical Computer*
744 *Science*, 14(2):195–208, 1981.
745 **33** Gabriel Thierrin. Permutation automata. *Theory of Computing Systems*, 2(1):83—-90, 1968.
746 **34** Wolfgang Thomas. Classifying regular events in symbolic logic. *Journal of Computer and*
747 *System Sciences*, 25(3):360–376, 1982.
748 **35** Bret Tilson. Categories as algebra: essential ingredient in the theory of monoids. *Journal of*
749 *Pure and Applied Algebra*, 48(1):83–198, 1987.
750 **36** Georg Zetzsche. Separability by piecewise testable languages and downward closures beyond
751 subwords. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer*
752 *Science*, LICS'18, pages 929–938, 2018.

Appendix

In this appendix, we present the proof of Theorem 22. Let us first recall the statement.

▶ **Theorem 22.** *Let $\mathcal{G}$ be a group prevariety and $\mathcal{A} = (Q, \delta)$ an NFA. Then, $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$ is the greatest $(BPol, +)$-sound subset of Q^4 for $\mathcal{G}$ and $\mathcal{A}$.*

The proof argument is based on the same outline as the one presented for Theorem 20 in the main paper. We fix a group prevariety $\mathcal{G}$ and an NFA $\mathcal{A} = (Q, \delta)$. Let $S \subseteq Q^4$ be the greatest $(BPol, +)$-sound subset for $\mathcal{G}$ and $\mathcal{A}$. We prove that $S = \mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$.

First part: $S \subseteq \mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$. We use *tuple separation* and Lemma 17. Let us start with terminology. For every $n \geq 1$ and $(q_1, r_1, q_2, r_2) \in Q^4$, we associate an n-tuple $T_n(q_1, r_1, q_2, r_2)$. We use induction on n and tuple concatenation to present the definition. If $n = 1$ then, $T_1(q_1, r_1, q_2, r_2) = (L_{\mathcal{A}}(q_2, r_2))$. If $n > 1$, then,

$$T_n(q_1, r_1, q_2, r_2) = \begin{cases} (L_{\mathcal{A}}(q_2, r_2)) \cdot T_{n-1}(q_1, r_1, q_2, r_2) & \text{if } n \text{ is odd} \\ (L_{\mathcal{A}}(q_1, r_1)) \cdot T_{n-1}(q_1, r_1, q_2, r_2) & \text{if } n \text{ is even.} \end{cases}$$

We use induction on n to prove the following proposition.

▶ **Proposition 29.** *For every $n \geq 1$ and $(q_1, r_1, q_2, r_2) \in S$, the n-tuple $T_n(q_1, r_1, q_2, r_2)$ is not $Pol(\mathcal{G}^+)$-separable under $\mathcal{G}$-control.*

By definition, Proposition 29 implies that for every $p \geq 1$ and every $(q_1, r_1, q_2, r_2) \in S$, the $2p$-tuple $(L_{\mathcal{A}}(q_1, r_1), L_{\mathcal{A}}(q_2, r_2))^p$ is not $Pol(\mathcal{G}^+)$-separable under $\mathcal{G}$-control. By Corollary 15, it follows that $L_{\mathcal{A}}(q_1, r_1)$ is not $BPol(\mathcal{G}^+)$-separable from $L_{\mathcal{A}}(q_2, r_2)$ under $\mathcal{G}$-control, *i.e.* that $(q_1, r_1, q_2, r_2) \in \mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$. We get $S \subseteq \mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}]$ as desired.

We prove Proposition 29 using induction on n. We fix $n \geq 1$ for the proof. In order to exploit the fact that S is $(BPol, +)$-sound, we need a property of the NFA $\mathcal{B}_S^+ = (Q^3, \gamma_S)$ used to define $\tau^+_{\mathcal{A},\mathcal{G}}$. When $n \geq 2$, this is where we use induction on n and Lemma 17.

▶ **Lemma 30.** *Consider $(s_1, s_2, s_3), (t_1, t_2, t_3) \in Q^3$ and a group language $H \subseteq A^*$. Assume that $H \cap L_{\mathcal{B}_S^+}((s_1, s_2, s_3), (t_1, t_2, t_3)) \neq \emptyset$. Then, $H \cap L_{\mathcal{A}}(s_1, t_1) \neq \emptyset$ and, if $n \geq 2$, then the n-tuple $(H \cap L_{\mathcal{A}}(s_1, t_1)) \cdot T_{n-1}(s_2, t_2, s_3, t_3)$ is not $Pol(\mathcal{G}^+)$-separable.*

Proof. By hypothesis, there exists $w \in H \cap L_{\mathcal{B}_S^+}((s_1, s_2, s_3), (t_1, t_2, t_3))$. Hence, the NFA $\mathcal{B}_S^+$ contains some run labeled by w from (s_1, s_2, s_3) to (t_1, t_2, t_3). We use a sub-induction on the number of transitions involved in that run. When no transitions are used: we have $w = \varepsilon$ and $(s_1, s_2, s_3) = (t_1, t_2, t_3)$. It follows that $w = \varepsilon \in H \cap L_{\mathcal{A}}(s_1, t_1)$. Moreover, if $n \geq 2$, the n-tuple $(H \cap L_{\mathcal{A}}(s_1, t_1)) \cdot T_{n-1}(s_2, s_2, s_3, s_3)$ is not $Pol(\mathcal{G}^+)$-separable by Lemma 13 since $\varepsilon \in L_{\mathcal{A}}(s_2, s_2) \cap L_{\mathcal{A}}(s_3, s_3)$. We now assume that at least one transition is used. We get a triple $(q_1, q_2, q_3) \in Q^3$, a word $w' \in A^*$ and $x \in A \cup \{\varepsilon\}$ such that we have $w = w'x$, $w' \in L_{\mathcal{B}_S^+}((s_1, s_2, s_3), (q_1, q_2, q_3))$ and $((q_1, q_2, q_3), x, (t_1, t_2, t_3)) \in \gamma_S^+$. Since H is a group language, it is recognized by a morphism $\alpha : A^* \to G$ into a finite group G. Let $H' = \alpha^{-1}(\alpha(w'))$. Clearly, H' is a group language and $w' \in H' \cap L_{\mathcal{B}_S^+}((s_1, s_2, s_3), (q_1, q_2, q_3))$. Thus, induction yields that $H' \cap L_{\mathcal{A}}(s_1, q_1) \neq \emptyset$ and, if $n \geq 2$, the n-tuple $(H' \cap L_{\mathcal{A}}(s_1, q_1)) \cdot T_{n-1}(s_2, q_2, s_3, q_3)$ is not $Pol(\mathcal{G}^+)$-separable. We now consider two cases depending on $x \in A \cup \{\varepsilon\}$.

Assume first that $x = a \in A$: we have $((q_1, q_2, q_3), a, (t_1, t_2, t_3)) \in \gamma_S^+$. By definition, it follows that $(q_i, a, t_i) \in \delta$ for $i = \{1, 2, 3\}$. Observe that $(H' \cap L_{\mathcal{A}}(s_1, q_1))a \subseteq H \cap L_{\mathcal{A}}(s_1, t_1)$. Indeed, if $u \in (H' \cap L_{\mathcal{A}}(s_1, q_1))a$, then $u = u'a$ where $u' \in H'$ and $u' \in L_{\mathcal{A}}(s_1, q_1)$. Since $H' = \alpha^{-1}(\alpha(w'))$, the hypothesis that $u' \in H'$ yields $\alpha(u) = \alpha(u'a) = \alpha(w'a) = \alpha(w)$ which

implies that $u \in H$ since $w \in H$ and H is recognized by α. Moreover, since $u' \in L_{\mathcal{A}}(s_1, q_1)$ and $(q_1, a, t_1) \in \delta$, we get $u = u'a \in L_{\mathcal{A}}(s_1, t_1)$. Altogether, this yields $u \in H \cap L_{\mathcal{A}}(s_1, t_1)$ as desired. Since we already know that $H' \cap L_{\mathcal{A}}(s_1, q_1) \neq \emptyset$, we get $H \cap L_{\mathcal{A}}(s_1, t_1) \neq \emptyset$. Moreover, if $n \geqslant 2$, since $(q_2, a, t_2), (q_3, a, t_3) \in \delta$, Lemma 13 yields that $(\{a\}) \cdot T_{n-1}(q_2, t_2, q_3, t_3)$ is not $Pol(\mathcal{G}^+)$-separable. Hence, since we already know that $(H' \cap L_{\mathcal{A}}(s_1, q_1)) \cdot T_{n-1}(s_2, q_2, s_3, q_3)$ is not $Pol(\mathcal{G}^+)$-separable and $(H' \cap L_{\mathcal{A}}(s_1, q_1))a \subseteq H \cap L_{\mathcal{A}}(s_1, t_1)$, it follows from Lemma 14 that $(H \cap L_{\mathcal{A}}(s_1, t_1)) \cdot T_{n-1}(s_2, t_2, s_3, t_3)$ is not $Pol(\mathcal{G}^+)$-separable.

Finally, assume that $x = \varepsilon$: we have $((q_1, q_2, q_3), \varepsilon, (t_1, t_2, t_3)) \in \gamma_S^+$. By definition, it follows that $q_1 = t_1$, $(q_2, t_2, q_3, t_3) \in S$ and there exists a nonempty word $y \in A^+$ which belongs to $L_{\mathcal{A}}(q_1, q_1)$, $L_{\mathcal{A}}(q_2, q_2)$, $L_{\mathcal{A}}(q_3, q_3)$, $L_{\mathcal{A}}(t_2, t_2)$ and $L_{\mathcal{A}}(t_3, t_3)$. Since $x = \varepsilon$, we have $w = w'$. Hence, since $w \in H$ and H is recognized by α, we obtain that $H' = \alpha(\alpha^{-1}(w')) \subseteq H$. Since $H' \cap L_{\mathcal{A}}(s_1, q_1) \neq \emptyset$ and $q_1 = t_1$, we get $H \cap L_{\mathcal{A}}(s_1, t_1) \neq \emptyset$. We now assume that $n \geqslant 2$. Since G is a finite group, there exists $k \geqslant 1$ such that $\alpha(y^k) = 1_G$. We write $z = y^k$. By hypothesis on y, we also have $z \in L_{\mathcal{A}}(q_1, q_1)$. It follows that $z^+ \subseteq \alpha^{-1}(1_G) \cap L_{\mathcal{A}}(q_1, q_1)$. Additionally, since z belongs to $L_{\mathcal{A}}(q_2, q_2)$, $L_{\mathcal{A}}(q_3, q_3)$, $L_{\mathcal{A}}(t_2, t_2)$ and $L_{\mathcal{A}}(t_3, t_3)$, we know that $z^+ L_{\mathcal{A}}(q_2, t_2) z^+ \subseteq L_{\mathcal{A}}(q_2, t_2)$ and $z^+ L_{\mathcal{A}}(q_3, t_3) z^+ \subseteq L_{\mathcal{A}}(q_3, t_3)$. Since $(q_2, t_2, q_3, t_3) \in S$, it follows from induction on n in Proposition 29 that the $(n-1)$-tuple $T_{n-1}(q_2, t_2, q_3, t_3)$ is not $Pol(\mathcal{G}^+)$-separable under $\mathcal{G}$-control. Altogether, we obtain from Lemma 17 that the n-tuple $(\alpha^{-1}(1_G) \cap L_{\mathcal{A}}(q_1, q_1)) \cdot T_{n-1}(q_2, t_2, q_3, t_3)$ is not $Pol(\mathcal{G}^+)$-separable. Finally, since $q_1 = t_1$ and $H' \subseteq H$, one may verify that $(H' \cap L_{\mathcal{A}}(s_1, q_1))(\alpha^{-1}(1_G) \cap L_{\mathcal{A}}(q_1, q_1)) \subseteq (H \cap L_{\mathcal{A}}(s_1, t_1))$. Since we already know that $(H' \cap L_{\mathcal{A}}(s_1, q_1)) \cdot T_{n-1}(s_2, q_2, s_3, q_3)$ is not $Pol(\mathcal{G}^+)$-separable, Lemma 14 yields that $(H \cap L_{\mathcal{A}}(s_1, t_1)) \cdot T_{n-1}(s_2, t_2, s_3, t_3)$ is not $Pol(\mathcal{G}^+)$-separable. ◀

We may now complete the proof of Proposition 29. By symmetry, we only treat the case when n is odd and leave the even case to the reader. Let $(q_1, r_1, q_2, r_2) \in S$, we have to prove that $T_n(q_1, r_1, q_2, r_2)$ is not $Pol(\mathcal{G}^+)$-separable under $\mathcal{G}$-control. Hence, we fix $H \in \mathcal{G}$ such that $\varepsilon \in H$ and prove $H \cap T_n(q_1, r_1, q_2, r_2)$ is not $Pol(\mathcal{G}^+)$-separable. Since S is $(BPol, +)$-sound, we have $\tau^+_{\mathcal{A},\mathcal{G}}(S) = S$ which implies that $(q_1, r_1, q_2, r_2) \in \tau^+_{\mathcal{A},\mathcal{G}}(S)$. Hence, it follows from (2) that $\{\varepsilon\}$ is not $\mathcal{G}$-separable from $L_{\mathcal{B}_S^+}((q_2, q_1, q_2), (r_2, r_1, r_2))$. Since $H \in \mathcal{G}$ and $\varepsilon \in H$, it follows that $H \cap L_{\mathcal{B}_S^+}((q_2, q_1, q_2), (r_2, r_1, r_2)) \neq \emptyset$. If $n = 1$, Lemma 30 yields $H \cap L_{\mathcal{A}}(q_2, r_2) \neq \emptyset$. Since $T_1(q_1, r_1, q_2, r_2) = (L_{\mathcal{A}}(q_2, r_2))$, we get that $H \cap T_1(q_1, r_1, q_2, r_2)$ is not $Pol(\mathcal{G}^+)$-separable as desired. If $n \geqslant 2$, then Lemma 30 implies that $(H \cap L_{\mathcal{A}}(s_1, t_1)) \cdot T_{n-1}(s_2, t_2, s_3, t_3)$ is not $Pol(\mathcal{G}^+)$-separable. Thus, since $H \in \mathcal{G} \subseteq Pol(\mathcal{G}^+)$, one may verify that the n-tuple $(H \cap L_{\mathcal{A}}(q_2, r_2)) \cdot (H \cap T_{n-1}(q_1, r_1, q_2, r_2))$ is not $Pol(\mathcal{G}^+)$-separable. By definition, this exactly says that $H \cap T_n(q_1, r_1, q_2, r_2)$ is not $Pol(\mathcal{G}^+)$-separable, completing the proof.

Second part: $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}] \subseteq S$. Consider an arbitrary set $R \subseteq Q^4$. We say that R is multiplication-closed to indicate that for every $(q, r, s, t) \in R$ and $(q', r', s', t') \in R$, if $r = q'$ and $t = s'$, then $(q, r', s, t') \in R$. Moreover, we say that an arbitrary set $R \subseteq Q^4$ is *good* if it is multiplication-closed and there are $L \in \mathcal{G}$ such $\varepsilon \in L$ and a $BPol(\mathcal{G}^+)$-cover $\mathbf{K}$ of L which is separating for R.

▶ **Proposition 31.** *Let $R \subseteq Q^4$. If R is good, then $\tau^+_{\mathcal{A},\mathcal{G}}(R)$ is good as well.*

We use Proposition 31 to complete the proof. Let $S_0 = Q^4$ and $S_i = \tau^+_{\mathcal{A},\mathcal{G}}(S_{i-1})$ for $i \geqslant 1$. By Lemma 21, we have $S_0 \supseteq S_1 \subseteq S_2 \supseteq \cdots$ and the is $n \in \mathbb{N}$ such that S_n is the greatest $(BPol, +)$-sound subset for $\mathcal{G}$ and $\mathcal{A}$, *i.e.* such that $S_n = S$. Since S_0 is good (it is clearly multiplication-closed and $\{A^*\}$ is a $BPol(\mathcal{G}^+)$-cover of $A^* \in \mathcal{G}$ which is separating for $S_0 = Q^4$), Proposition 31 implies that S_i is good for all $i \in \mathbb{N}$. Hence, $S = S_n$ is good. We

get $L \in \mathcal{G}$ such $\varepsilon \in L$ and a $BPol(\mathcal{G}^+)$-cover $\mathbf{K}$ of L which is separating for S. By Lemma 11, this yields $\mathcal{I}_{BPol(\mathcal{G}^+)}[\mathcal{G}, \mathcal{A}] \subseteq S$ as desired.

We turn to Proposition 25. Let $R \subseteq Q^4$ be a good set. We have to prove that $\tau^+_{\mathcal{A},\mathcal{G}}(R)$ is multiplication-closed and build $L \in \mathcal{G}$ such $\varepsilon \in L$ and a $BPol(\mathcal{G}^+)$-cover $\mathbf{K}$ of L which is separating for $\tau^+_{\mathcal{A},\mathcal{G}}(R)$. This proves that $\tau^+_{\mathcal{A},\mathcal{G}}(R)$ is good as desired. Let us first prove that $\tau^+_{\mathcal{A},\mathcal{G}}(R)$ is multiplication-closed (we use the hypothesis that R is good).

▶ **Lemma 32.** *The set $\tau^+_{\mathcal{A},\mathcal{G}}(R) \subseteq Q^4$ is multiplication-closed.*

Proof. Let $(q,r,s,t) \in \tau^+_{\mathcal{A},\mathcal{G}}(R)$ and $(q',r',s',t') \in \tau^+_{\mathcal{A},\mathcal{G}}(R)$ such that $r = q'$ and $t = s'$. We need to prove that $(q,r',s,t') \in \tau^+_{\mathcal{A},\mathcal{G}}(R)$. By (2) in the definition, this boils down to proving that $\{\varepsilon\}$ is *not* $\mathcal{G}$-separable from $L_{\mathcal{B}^+_R}((s,q,s),(t',r',t'))$ and $L_{\mathcal{B}^+_R}((q,s,q),(r',t',r'))$. By symmetry, we only prove the former. By hypothesis on (q,r,s,t) and (q',r',s',t'), we get from (2) that $\{\varepsilon\}$ is *not* $\mathcal{G}$-separable from both $L_{\mathcal{B}^+_R}((s,q,s),(t,r,t))$ and $L_{\mathcal{B}^+_R}((s',q',s'),(t',r',t'))$. Since $\mathcal{G}$ is a prevariety it then follows from Lemma 14 that $\{\varepsilon\}$ is not $\mathcal{G}$-separable from the concatenation $L_{\mathcal{B}^+_R}((s,q,s),(t,r,t))L_{\mathcal{B}^+_R}((s',q',s'),(t',r',t'))$. Finally, since $(t,r,t) = (s',q',s')$, we know that $L_{\mathcal{B}^+_R}((s,q,s),(t,r,t))L_{\mathcal{B}^+_R}((s',q',s'),(t',r',t')) \subseteq L_{\mathcal{B}^+_R}((s,q,s),(t',r',t'))$. We conclude that $\{\varepsilon\}$ is *not* $\mathcal{G}$-separable from both $L_{\mathcal{B}^+_R}((s,q,s),(t',r',t'))$ as desired. ◀

We now build $L \in \mathcal{G}$ such that $\varepsilon \in L$ (this part is independent from our hypothesis on R).

▶ **Lemma 33.** *There exists $L \in \mathcal{G}$ such that $\varepsilon \in L$ and for every $(q,r,s,t) \in Q^4$, if $L_{\mathcal{B}^+_R}((q,s,q),(r,t,r)) \cap L \neq \emptyset$ and $L_{\mathcal{B}^+_R}((s,q,s),(t,r,t)) \cap L \neq \emptyset$, then $(q,r,s,t) \in \tau^+_{\mathcal{A},\mathcal{G}}(R)$.*

Proof. Let $\mathbf{H}$ be the *finite* set of all languages recognized by $\mathcal{B}^+_R$ such that $\{\varepsilon\}$ is $\mathcal{G}$-separable from H. For every $H \in \mathbf{H}$, there exists $L_H \in \mathcal{G}$ such that $\varepsilon \in L_H$ and $L_H \cap H = \emptyset$. We define $L = \bigcap_{H \in \mathbf{H}} L_H \in \mathcal{G}$. It is clear that $\varepsilon \in L$. Moreover, given $(q,r,s,t) \in Q^4$, if $L_{\mathcal{B}^+_R}((q,s,q),(r,t,r)) \cap L \neq \emptyset$ and $L_{\mathcal{B}^+_R}((s,q,s),(t,r,t)) \cap L \neq \emptyset$, it follows from the definition of L that $\{\varepsilon\}$ is not $\mathcal{G}$-separable from both $L_{\mathcal{B}^+_R}((q,s,q),(r,t,r))$ and $L_{\mathcal{B}^+_R}((s,q,s),(t,r,t))$. It then follows from (2) in the definition of $\tau^+_{\mathcal{A},\mathcal{G}}$ that $(q,r,s,t) \in \tau^+_{\mathcal{A},\mathcal{G}}(R)$. ◀

We fix $L \in \mathcal{G}$ as described in Lemma 33 for the remainder of the proof. We now build the $BPol(\mathcal{G}^+)$-cover $\mathbf{K}$ of L using the hypothesis that R is good and Proposition 7.

▶ **Lemma 34.** *For all $(q,r) \in Q^2$, there is $H_{q,r} \in BPol(\mathcal{G}^+)$ such that $L_{\mathcal{A}}(q,r) \cap L \subseteq H_{q,r}$ and for all pairs $(s,t) \in Q^2$, if $L_{\mathcal{A}}(s,t) \cap H_{q,r} \neq \emptyset$ then $L_{\mathcal{B}^+_R}((q,s,q),(r,t,r)) \cap L \neq \emptyset$.*

Proof. Since R is good, there are $U \in \mathcal{G}$ such that $\varepsilon \in U$ and a $BPol(\mathcal{G}^+)$-cover $\mathbf{V}$ of U which is separating for R. We use them to build $H_{q,r}$. Since $U \in \mathcal{G}$ and $\varepsilon \in U$ Proposition 7 yields a cover $\mathbf{P}$ of $L_{\mathcal{A}}(q,r) \cap L$ such that for each $P \in \mathbf{P}$, there exists a word $w_P \in L_{\mathcal{A}}(q,r) \cap L$ and an $\mathcal{A}$-guarded decomposition $(w_1, \ldots, w_{n+1})$ of w_P for some $n \in \mathbb{N}$ such that $P = w_1 U \cdots w_n U w_{n+1}$ (if $n = 0$, then $P = \{w_1\}$). Now, for every $P \in \mathbf{P}$, we build a $BPol(\mathcal{G}^+)$-cover $\mathbf{K}_P$ of P from the cover $\mathbf{V}$ of U. Let $(w_1, \ldots, w_{n+1})$ be the $\mathcal{A}$-guarded decomposition of w_P such that $P = w_1 U \cdots w_n U w_{n+1}$ (in particular, this means that P is of the form $U_0 a_1 U_1 \cdots a_m U_m$ where $a_1 \cdots a_m = w_1 \cdots w_n$ and $U_i = U$ or $U_i = \{\varepsilon\}$ for each $i \leq m$). By definition, $\mathbf{V}$ is a $BPol(\mathcal{G}^+)$-cover of $U \in \mathcal{G} \subseteq Pol(\mathcal{G}^+)$. Moreover, we have $\{\varepsilon\} \in \mathcal{G}^+ \subseteq Pol(\mathcal{G}^+)$ by definition of $\mathcal{G}^+$ and $\{\{\varepsilon\}\}$ is a $BPol(\mathcal{G}^+)$-cover of $\{\varepsilon\}$. Hence, Proposition 5 yields a $BPol(\mathcal{G}^+)$-cover $\mathbf{K}_P$ of $P = w_1 U \cdots w_n U w_{n+1}$ such that for every $K \in \mathbf{K}_P$, there exist $V_1, \ldots, V_n \in \mathbf{V}$ such that $K \subseteq w_1 V_1 \cdots w_n V_n w_{n+1}$. We define $H_{q,r}$ as the union of all languages K such that $K \in \mathbf{K}_P$ for some $P \in \mathbf{P}$ and $L_{\mathcal{A}}(q,r) \cap K \neq \emptyset$. Clearly, $H_{q,r} \in BPol(\mathcal{G}^+)$. Moreover, since $\mathbf{P}$ is a cover of $L_{\mathcal{A}}(q,r) \cap L$, and $\mathbf{K}_P$ is a cover

of P for each $P \in \mathbf{P}$, it is clear that $L_{\mathcal{A}}(q,r) \cap L \subseteq H_{q,r}$. We now fix $(s,t) \in Q^2$ such that $L_{\mathcal{A}}(s,t) \cap H_{q,r} \neq \emptyset$ and show that $L_{\mathcal{B}_R^+}((q,s,q),(r,t,r)) \cap L \neq \emptyset$. By definition of $H_{q,r}$, we get $P \in \mathbf{P}$ and $K \in \mathbf{K}_P$ such that $L_{\mathcal{A}}(q,r) \cap K \neq \emptyset$ and $L_{\mathcal{A}}(s,t) \cap K \neq \emptyset$. By definition, $P = w_1 U \cdots w_n U w_{n+1}$ where $(w_1, \ldots, w_{n+1})$ is an $\mathcal{A}$-guarded decomposition of $w_P \in L_{\mathcal{A}}(q,r) \cap L$. We use w_P to build a new word $w' \in L_{\mathcal{B}_R^+}((q,s,q),(r,t,r)) \cap L$.

We fix $x \in L_{\mathcal{A}}(s,t) \cap K$ and $y \in L_{\mathcal{A}}(q,r) \cap K$. Since $w_P = w_1 \cdots w_{n+1}$ and $w_P \in L_{\mathcal{A}}(q,r)$, we may decompose the corresponding run in $\mathcal{A}$: we get $p_0, \ldots, p_{n+1} \in Q$ such that $p_0 = q$, $p_{n+1} = r$ and $w_i \in L_{\mathcal{A}}(p_{i-1}, p_i)$ for $1 \leq i \leq n+1$. Moreover, since $K \in \mathbf{K}_P$, we have $K \subseteq w_1 V_1 \cdots w_n V_n w_{n+1}$ for $V_1, \ldots, V_n \in \mathbf{V}$ (if $n = 0$, then $K \subseteq \{w_1\}$). Since $x, y \in K$, we get $x_i, y_i \in V_i$ for $1 \leq i \leq n$ such that $x = w_1 x_1 \cdots w_n x_n w_{n+1}$ and $y = w_1 y_1 \cdots w_n y_n w_{n+1}$. Since $x \in L_{\mathcal{A}}(s,t)$, we get $s_1, t_1, \ldots, s_{n+1}, t_{n+1} \in Q$ where $s_1 = s$, $t_{n+1} = t$, $w_i \in L_{\mathcal{A}}(s_i, t_i)$ for $1 \leq i \leq n+1$ and $x_i \in L_{\mathcal{A}}(t_i, s_{i+1})$ for $1 \leq i \leq n$. Symmetrically, since $y \in L_{\mathcal{A}}(q,r)$, we get $q_1, r_1, \ldots, q_{n+1}, r_{n+1} \in Q$ with $q_1 = q$, $r_{n+1} = r$, $w_i \in L_{\mathcal{A}}(q_i, r_i)$ for $1 \leq i \leq n+1$, and $y_i \in L_{\mathcal{A}}(r_i, q_{i+1})$ for $1 \leq i \leq n$. First, note that when $n = 0$, we have $w_P = w_1$ and the above implies that $w_P \in L_{\mathcal{A}}(q,r)$ and $w_P \in L_{\mathcal{A}}(s,t)$. Thus, $w_P \in L_{\mathcal{B}_R^+}((q,s,q),(r,t,r))$ by definition of the labeled transition in $\mathcal{B}_R^+$. This concludes the proof since we also know that $w_P \in L$. We now assume that $n \geq 1$.

By hypothesis, $(w_1, \ldots, w_{n+1})$ is an $\mathcal{A}$-guarded decomposition. Hence, for $1 \leq i \leq n$, we get $z_i \in A^+$ which is a right $\mathcal{A}$-loop for w_i and a left $\mathcal{A}$-loop for w_{i+1}. Let $\alpha : A^* \to G$ be a morphism into a finite group G recognizing both L and U (recall that L and U are group languages). Since g is a finite group, there exists $k \geq 1$ such that for each $1 \leq i \leq n$, we have $\alpha(z_i^k) = 1_G$. We let $u_i = z_i^k$ for $1 \leq i \leq n$. One may verify that u_i remains a right $\mathcal{A}$-loop for w_i and a left $\mathcal{A}$-loop for w_{i+1}. Moreover, since $\alpha(u_i) = 1_G$, we know that $u_i \in U$ (recall that $\varepsilon \in U$ and U is recognized by α). We let $w'_1 = w_1 u_1$, $w'_{n+1} = u_n w_{n+1}$ and $w'_i = u_{i-1} w_i u_i$ for $2 \leq i \leq n$. Finally, we let $w' = w'_1 \cdots w'_n w'_{n+1}$ and show that $w' \in L \cap L_{\mathcal{B}_R^+}((q,s,q),(r,t,r))$ which completes the proof. First, since $\alpha(u_i) = 1_G$ for $1 \leq i \leq n$, it is immediate that $\alpha(w') = \alpha(w_1 \cdots w_n w_{n+1}) = \alpha(w_P)$. Since $w_P \in L$ which is recognized by α, we get $w' \in L$.

We now concentrate on proving that $w' \in L_{\mathcal{B}_R^+}((q,s,q),(r,t,r))$. For $1 \leq i \leq n+1$, we know that w_i belongs to $L_{\mathcal{A}}(p_{i-1}, p_i)$, $L_{\mathcal{A}}(s_i, t_i)$ and $L_{\mathcal{A}}(q_i, r_i)$. Hence, one may verify from the definition of left/right $\mathcal{A}$-loops that there are $p'_0, \ldots, p'_{n+1} \in Q$, $s'_1, t'_1, \ldots, s'_{n+1}, t'_{n+1} \in Q$ and $q'_1, r'_1, \ldots, q'_{n+1}, r'_{n+1} \in Q$ such that,

- $p'_0 = p_0 = q$, $p'_{n+1} = p_{n+1} = r$, $w'_i \in L_{\mathcal{A}}(p'_{i-1}, p'_i)$ for $1 \leq i \leq n+1$ and $u_i \in L_{\mathcal{A}}(p'_i, p'_i)$ for $1 \leq i \leq n$.
- $s'_0 = s_0 = s$, $t'_{n+1} = t_{n+1} = t$, $w'_i \in L_{\mathcal{A}}(s'_i, t'_i)$ for $1 \leq i \leq n+1$ and we have $u_i \in L_{\mathcal{A}}(t'_i, t'_i) \cap L_{\mathcal{A}}(t'_i, t_i) \cap L_{\mathcal{A}}(s_{i+1}, s'_{i+1}) \cap L_{\mathcal{A}}(s'_{i+1}, s'_{i+1})$ for $1 \leq i \leq n$.
- $q'_0 = q_0 = q$, $r'_{n+1} = r_{n+1} = r$, $w'_i \in L_{\mathcal{A}}(q'_i, r'_i)$ for $1 \leq i \leq n+1$ and we have $u_i \in L_{\mathcal{A}}(r'_i, r'_i) \cap L_{\mathcal{A}}(r'_i, r_i) \cap L_{\mathcal{A}}(q_{i+1}, q'_{i+1}) \cap L_{\mathcal{A}}(q'_{i+1}, q'_{i+1})$ for $1 \leq i \leq n$.

By definition of the labeled transitions in the NFA $\mathcal{B}_R^+$, it is straightforward to verify that we have $w'_i \in L_{\mathcal{B}_R^+}((p'_{i-1}, s'_i, q'_i), (p'_i, t'_i, r'_i))$ for $1 \leq i \leq n+1$. We now prove the following fact.

▶ **Fact 35.** *For $1 \leq i \leq n$, we have $((p'_i, t'_i, r'_i), \varepsilon, (p'_i, s'_{i+1}, q'_{i+1})) \in \gamma_R^+$.*

Proof. We fix i for the proof. Since we know that $u_i \in A^+$ belongs to $L_{\mathcal{A}}(p'_i, p'_i)$, $L_{\mathcal{A}}(t'_i, t'_i)$, $L_{\mathcal{A}}(r'_i, r'_i)$, $L_{\mathcal{A}}(s'_{i+1}, s'_{i+1})$ and $L_{\mathcal{A}}(q'_{i+1}, q'_{i+1})$, it suffices to prove that $(t'_i, s'_{i+1}, r'_i, q'_{i+1}) \in R$. This will imply that $((p'_i, t'_i, r'_i), \varepsilon, (p'_i, s'_{i+1}, q'_{i+1})) \in \gamma_R^+$ by definition of γ_R^+. Recall that $x_i \in L_{\mathcal{A}}(t_i, s_{i+1})$, $y_i \in L_{\mathcal{A}}(r_i, q_{i+1})$ and $x_i, y_i \in V_i$. Since $V_i \in \mathbf{V}$ which is *separating* for R, it follows that $(t_i, s_{i+1}, r_i, q_{i+1}) \in R$. Moreover, $u_i \in U$ which yields $V \in \mathbf{V}$ such that $u_i \in V$ since $\mathbf{V}$ is a cover of U. Hence, since $u_i \in L_{\mathcal{A}}(t'_i, t_i)$ and $u_i \in L_{\mathcal{A}}(r'_i, r_i)$. The hypothesis that $\mathbf{V}$ is separating for R also yields $(t'_i, t_i, r'_i, r_i) \in R$. Symmetrically, one may use the hypotheses

that $u_i \in L_{\mathcal{A}}(s_{i+1}, s'_{i+1})$ and $u_i \in L_{\mathcal{A}}(q_{i+1}, q'_{i+1})$ to verify that $(s_{i+1}, s'_{i+1}, q_{i+1}, q'_{i+1}) \in R$. Altogether, since R is multiplication-closed, we get $(t'_i, s'_{i+1}, r'_i, q'_{i+1}) \in R$ as desired. ◀

In view of Fact 35, we obtain $w' = w'_1 \cdots w'_n w'_{n+1} \in L_{\mathcal{B}_R^+}((p'_0, s'_1, q'_1), (p'_{n+1}, t'_{n+1}, r'_{n+1}))$. This exactly says that $w' \in L_{\mathcal{B}_R^+}((q, s, q), (r, t, r))$ which completes the proof. ◀

We may now build $\mathbf{K}$. Let $\mathbf{H} = \{H_{q,r} \mid (q, r) \in Q^2\}$. Consider the following equivalence $\sim$ defined on L: given $u, v \in L$, we let $u \sim v$ if and only if $u \in H_{q,r} \Leftrightarrow v \in H_{q,r}$ for every $(q, r) \in Q^2$. We let $\mathbf{K}$ as the partition of L into $\sim$-classes. Clearly, each $K \in \mathbf{K}$ is a Boolean combination involving the languages in $\mathbf{H}$ (which belong to $BPol(\mathcal{G}^+)$) and $L \in \mathcal{G}$. Hence, $\mathbf{K}$ is a $BPol(\mathcal{G}^+)$-cover of L. It remains to prove that it is separating for $\tau^+_{\mathcal{A},\mathcal{G}}(R)$. Let $q, r, s, t \in Q$ and $K \in \mathbf{K}$ such that there are $u \in L_{\mathcal{A}}(q, r) \cap K$ and $v \in L_{\mathcal{A}}(s, t) \cap K$. By definition of $\mathbf{K}$, we have $u, v \in L$ and $u \sim v$. In particular, we have $u \in L_{\mathcal{A}}(q, r) \cap L$ which yields $u \in H_{q,r}$ by definition in Lemma 34. Together with $u \sim v$, this yields $v \in H_{q,r}$. Hence, $L_{\mathcal{A}}(s, t) \cap H_{q,r} \neq \emptyset$ and Lemma 34 yields $L_{\mathcal{B}_R^+}((q, s, q), (r, t, r)) \cap L \neq \emptyset$. One may now use a symmetrical argument to obtain $L_{\mathcal{B}_R^+}((s, q, s), (t, r, t)) \cap L \neq \emptyset$. By definition of L in Lemma 33, this yields $(q, r, s, t) \in \tau_{\mathcal{A},\mathcal{G}}(R)$, completing the proof.