Concatenation Hierarchies and Forbidden Patterns

Christian Glaßer[*] *Heinz Schmitz*[**]
Theoretische Informatik
Julius–Maximilians–Universität Würzburg
97074 Würzburg, Germany
{glasser, schmitz}@informatik.uni-wuerzburg.de

Abstract

We make the following progress on the dot–depth problem:

(1) We introduce classes $\mathcal{C}_n^{\mathcal{B}}$ and $\mathcal{C}_n^{\mathcal{L}}$ of starfree languages defined via forbidden patterns in finite automata. It is shown for all $n \geq 0$ that $\mathcal{C}_n^{\mathcal{B}}$ ($\mathcal{C}_n^{\mathcal{L}}$) contains level $n + 1/2$ of the dot–depth hierarchy (Straubing–Thérien hierarchy, resp.). Since we prove that $\mathcal{C}_n^{\mathcal{B}}$ and $\mathcal{C}_n^{\mathcal{L}}$ have decidable membership problems, this yields a lower bound algorithm for the dot–depth of a given language.

(2) We prove many structural similarities between our hierarchies $\{\mathcal{C}_n^{\mathcal{B}}\}$ and $\{\mathcal{C}_n^{\mathcal{L}}\}$, and the mentioned concatenation hierarchies. Both show the same inclusion structure and can be separated by the same languages. Moreover, we see that our pattern classes are not too large, since $\mathcal{C}_n^{\mathcal{L}}$ does not capture level $n + 1/2$ of the dot–depth hierarchy. We establish an effective conjecture for the dot–depth problem, namely that $\mathcal{C}_n^{\mathcal{B}}$ and $\mathcal{C}_n^{\mathcal{L}}$, and the respective levels of concatenation hierarchies in fact coincide. It is known from literature that this is true for $n = 0, 1$.

(3) We prove the decidability of level $5/2$ of the Straubing–Thérien hierarchy in case of a two–letter alphabet — in exactly the way as predicted by our conjecture. To our knowledge, no results concerning the decidability of this class (and of any level $\geq 5/2$ of these concatenation hierarchy) were previously known. The result is based on a more general upward translation of decidability, connecting concatenation hierarchies and our pattern hierarchies.

Keywords: Dot–depth problem, concatenation hierarchies, finite automata, decidability

[*] Supported by the Studienstiftung des Deutschen Volkes.
[**] Supported by the Deutsche Forschungsgemeinschaft (DFG), grant Wa 847/4-1.

Contents

1 Introduction **4**

2 Preliminaries **6**
2.1 Basic Definitions 6
2.2 Definition of Concatenation Hierarchies 8
2.3 Alternative Definitions 9
2.3.1 Polynomial Closure 9
2.3.2 Boolean Closure 11
2.3.3 Comparing the Lower Levels 12
2.3.4 Comparing the Hierarchies 13
2.4 Normalforms and Closure Properties 15

3 A Theory of Forbidden Patterns **18**
3.1 Definition of Iterated Patterns 18
3.2 Auxiliary Results for Iterated Patterns 19
3.3 Pattern Iterator versus Polynomial Closure 25
3.4 Inclusion Structure of the classes $\mathcal{C}_n^{\mathcal{T}}$ 27
3.5 Pattern Iteration remains Starfree 28
3.6 Decidability of the Pattern Classes 32

4 Consequences for Concatenation Hierarchies **38**
4.1 Some Auxiliary Results 38
4.2 Inclusion Relations and Lower Levels 38
4.3 Pattern Classes are Starfree 42
4.4 The Hierarchy of Pattern Classes is Strict 42
4.5 Decidability of the Pattern Classes $\mathcal{C}_n^{\mathcal{B}}$ and $\mathcal{C}_n^{\mathcal{L}}$ 44
4.6 Lower Bounds and a Conjecture for Concatenation Hierarchies 45

5 $\mathcal{L}_{5/2}$ is decidable for two–letter alphabets **47**
5.1 Changing the Alphabet 47
5.2 Automata Construction 49
5.3 Transformation of Patterns 51
5.4 Transformation of Expressions 54
5.5 Connecting Hierarchies 59

References **60**

List of Figures

1 Example for an appearance of a pattern at some state 35
2 Example for a connection of two states via a pattern 35
3 Concatenation Hierarchies and Forbidden Pattern Classes 45
4 Forbidden Pattern for $\mathcal{L}_{1/2}$ 62
5 Forbidden Pattern for $\mathcal{L}_{3/2}$ 62
6 Forbidden Pattern for $\mathcal{C}_2^{\mathcal{L}}$ 63
7 Forbidden Pattern for $\mathcal{B}_{1/2}$ 63
8 Forbidden Pattern for $\mathcal{B}_{3/2}$ 64
9 Forbidden Pattern for $\mathcal{C}_2^{\mathcal{B}}$ 64
10 Forbidden Pattern for $\mathcal{B}_{3/2}$ from [GS00] 65

1 Introduction

We contribute to the study of concatenation hierarchies and focus on starfree regular languages, which are constructed from alphabet letters using Boolean operations together with concatenation. Alternating these two kinds of operations leads to the definition of hierarchies that further classify starfree languages. Here we deal with the dot–depth hierarchy [CB71] and the closely related Straubing–Thérien hierarchy [Str81, Thé81, Str85]. Their investigation is of major interest in many research areas since surprisingly close connections have been exposed, e.g., to finite model theory, theory of finite semigroups, complexity theory and others (for an overview, see [Pin96]).

Let A be some finite alphabet with $|A| \geq 2$. For a class $\mathcal{C}$ of languages let $\mathrm{Pol}(\mathcal{C})$ be its polynomial closure, i.e., the closure under finite union and concatenation. Denote by $\mathrm{BC}(\mathcal{C})$ its Boolean closure (taking complements with respect to A^+). Then the classes $\mathcal{B}_{n/2}$ of the dot–depth hierarchy and the classes $\mathcal{L}_{n/2}$ of the Straubing–Thérien hierarchy can be defined as follows.

$$
\begin{array}{lll}
\mathcal{B}_{1/2} =_{\mathrm{def}} \mathrm{Pol}(\{\{w\} : w \in A^+\} \cup \{A^+\}) & \mathcal{L}_{1/2} =_{\mathrm{def}} \mathrm{Pol}(\{A^*aA^* : a \in A\}) & \\
\mathcal{B}_{n+1} =_{\mathrm{def}} \mathrm{BC}(\mathcal{B}_{n+1/2}) & \mathcal{L}_{n+1} =_{\mathrm{def}} \mathrm{BC}(\mathcal{L}_{n+1/2}) & \text{for } n \geq 0 \\
\mathcal{B}_{n+3/2} =_{\mathrm{def}} \mathrm{Pol}(\mathcal{B}_{n+1}) & \mathcal{L}_{n+3/2} =_{\mathrm{def}} \mathrm{Pol}(\mathcal{L}_{n+1}) & \text{for } n \geq 0
\end{array}
$$

By definition, all these classes are closed under union and it is known, that they are also closed under intersection [Arf91].

The question whether there exists an algorithm solving the membership problem for the levels of these hierarchies is known as the dot–depth problem, recently considered as one of the most important open questions on regular languages [Pin98]. Although many researchers believe the answer should be positive, some suspect the contrary. Up to now, levels $1/2$, 1 and $3/2$ of both hierarchies are known to be decidable [Sim75, Kna83, Arf87, PW97, GS00], while the question is open for any other level. Partial results are known only for level 2 of the Straubing–Thérien hierarchy. It is decidable if a two–letter alphabet is considered [Str88].

We take up the discussion started in [GS00] and continue known characterizations of lower levels in a natural way to a more general approach. More precisely, based on recent forbidden pattern characterizations of $\mathcal{L}_{1/2}$, $\mathcal{L}_{3/2}$ and $\mathcal{B}_{1/2}$ in [PW97], and of $\mathcal{B}_{3/2}$ in [GS00], we introduce hierarchies defined by forbidden patterns in deterministic finite automata (dfa, for short). We obtain strict and decidable hierarchies of classes $\mathcal{C}_n^{\mathcal{B}}$ and $\mathcal{C}_n^{\mathcal{L}}$, which exhaust the class of starfree languages and which are closely related to the dot–depth hierarchy and the Straubing–Thérien hierarchy. We prove that

$$
\begin{array}{lll}
\mathcal{B}_{1/2} = \mathcal{C}_0^{\mathcal{B}} & \mathcal{L}_{1/2} = \mathcal{C}_0^{\mathcal{L}} & \\
\mathcal{B}_{3/2} = \mathcal{C}_1^{\mathcal{B}} & \mathcal{L}_{3/2} = \mathcal{C}_1^{\mathcal{L}} & \\
\mathcal{B}_{n+1/2} \subseteq \mathcal{C}_n^{\mathcal{B}} & \mathcal{L}_{n+1/2} \subseteq \mathcal{C}_n^{\mathcal{L}} & \text{for } n \geq 2.
\end{array}
$$

These inclusions imply a lower bound algorithm for the dot–depth of a given language L. Just determine the class $\mathcal{C}_n^{\mathcal{B}}$ or $\mathcal{C}_n^{\mathcal{L}}$ for minimal n to which L belongs. It follows that L has at least dot–depth $n + 1/2$. For another result of this type see [Wei93].

We obtain more structural similarities. The pattern hierarchies show the same inclusion structure as known for concatenation hierarchies. Typical languages separating the levels of the dot–depth hierarchy witness also the strictness of our pattern hierarchies, and it holds that $\mathcal{C}_n^{\mathcal{L}}$ does not capture $\mathcal{B}_{n+1/2}$. We take all this as reasons to conjecture that in fact $\mathcal{B}_{n+1/2} = \mathcal{C}_n^{\mathcal{B}}$ and $\mathcal{L}_{n+1/2} = \mathcal{C}_n^{\mathcal{L}}$ for all $n \geq 0$. This conjecture agrees with the known decidability results, and we uniformly iterate with the definition of the pattern classes, what we have learned from these levels. If an effective characterization in terms of forbidden patterns in automata is possible at all, we believe that our conjecture provides a step in this direction. Other conjectures for the dot–depth problem can be found in the literature, e.g., the one in [PW97], note that we would have as a consequence an effective characterization.

Our results are applications of more general ones. For a so–called initial pattern $\mathcal{T}$ (fulfilling some reasonable weak prerequisites) we define $\mathcal{C}_0^{\mathcal{T}}$ to be the class of languages defined by $\mathcal{T}$ in a forbidden pattern manner, and we recursively iterate $\mathcal{T}$ to obtain $\mathcal{C}_n^{\mathcal{T}}$. Then we prove that $\mathcal{C}_n^{\mathcal{T}} \cup \mathrm{co}\mathcal{C}_n^{\mathcal{T}} \subseteq \mathcal{C}_{n+1}^{\mathcal{T}} \cap \mathrm{co}\mathcal{C}_{n+1}^{\mathcal{T}}$ and that $\mathrm{Pol}(\mathrm{co}\mathcal{C}_n^{\mathcal{T}}) \subseteq \mathcal{C}_{n+1}^{\mathcal{T}}$ hold. If one could also show $\mathrm{Pol}(\mathrm{co}\mathcal{C}_n^{\mathcal{T}}) \supseteq \mathcal{C}_{n+1}^{\mathcal{T}}$ then our conjecture would follow.

Finally, we give another argument in favour of the conjecture, i.e., we show that it is true in one more case, namely that $\mathcal{L}_{5/2} = \mathcal{C}_2^{\mathcal{L}}$ holds if we consider a two–letter alphabet. This implies the decidability of $\mathcal{L}_{5/2}$ for $|A| = 2$. To our knowledge, nothing was previously known concerning the decidability of $\mathcal{L}_{5/2}$ (nor for any other level $\geq 5/2$). It has immediate consequences in first–order logic. The connection of the Straubing–Thérien hierarchy to this logic goes back to the work of McNaughton and Papert [MP71]. It is related to the first–order logic $\mathrm{FO}[<]$ having unary relations for the alphabet symbols from A and the binary relation $<$. Let Σ_n be the subclass of this logic which is defined by at most $n-1$ quantifier alternations, starting with an existential quantifier. It has been proved in [Tho82, PP86] that Σ_n–formulas describe just the $\mathcal{L}_{n-1/2}$ languages. Due to this characterization we can conclude the decidability of the class of languages over $\{0,1\}^+$ definable by Σ_3–formulas of the logic $\mathrm{FO}[<]$.

The papers starts in section 2 with a discussion of various notions of polynomial closure and of dot–depth. It is shown in detail that if we define all hierarchy classes as classes of languages of A^+, as we did here, then the languages on each level differ from other definitions in the literature at most by the empty word (see Theorem 2.20). This is easy to handle when dealing with automata. The benefits we obtain hereby are on one hand the levelwise comparability by inclusion between the concatenation hierarchies (see Fig. 3), on the other hand we see that we can use the same operations to build up the dot–depth hierarchy and the Straubing–Thérien hierarchy. This is reflected on the pattern side by the uniform iteration rule in both cases. We just start with two different initial patterns.

2 Preliminaries

2.1 Basic Definitions

We fix some arbitrary finite alphabet A with $|A| \geq 2$. All definitions of language classes will be made with respect to A and it will always be clear from the context if we deal with a particular alphabet. The empty word is denoted by ε, the set of all words over A is denoted by A^* and the set of all non–empty words over A is denoted by A^+. We take complements with respect to A^* or A^+, depending on what domain we consider. So for a class $\mathcal{C}$ of languages of A^+ denote $\mathrm{co}^+\mathcal{C} =_{\mathrm{def}} \{ A^+ \setminus L \mid L \in \mathcal{C} \}$ the set of complements w.r.t. A^+, and for a class $\mathcal{C}$ of languages of A^* let $\mathrm{co}^*\mathcal{C} =_{\mathrm{def}} \{ A^* \setminus L \mid L \in \mathcal{C} \}$ be set of complements w.r.t. A^*. Since we focus on A^+ we write $\mathrm{co}\mathcal{C}$ instead of $\mathrm{co}^+\mathcal{C}$.

For a word $w \in A^*$ denote by $|w|$ its number of letters, and by $\alpha(w)$ the set of letters occurring in w. For $w \in A^*$ and $L \subseteq A^*$ we define the left and right residuals of L as $w^{-1}L =_{\mathrm{def}} \{ v \in A^* \mid wv \in L \}$ and $Lw^{-1} =_{\mathrm{def}} \{ v \in A^* \mid vw \in L \}$. We write $\mathcal{P}(B)$ for the powerset of an arbitrary set B.

Regular languages are build up from the empty set and the singletons $\{a_i\}$ for $a_i \in A$ using Boolean operations, concatenation and iteration. Of particular interest for us is the subclass of starfree languages, denoted as SF. Here the iteration operation is not allowed and we look at languages of A^+ (taking complements with respect to A^+).

A deterministic finite automaton (dfa) F is given by $F = (A, S, \delta, s_0, S')$, where A is its input alphabet, S is its set of states, $\delta : A \times S \to S$ is its total transition function, $s_0 \in S$ is the starting state and $S' \subseteq S$ is the set of accepting states. For a dfa F we denote by $L(F)$ the language accepted by F. As usual, we extend transition functions to input words, and we denote by $|F|$ the number of states of F. In this paper we will look only at dfa's where the starting state is not accepting. We say that a state $s \in S$ has a loop $v \in A^*$ (has a v-loop, for short) if and only if $\delta(s, v) = s$. Every $w \in A^*$ induces a total mapping $\delta^w : S \to S$ with $\delta^w(s) =_{\mathrm{def}} \delta(s, w)$. We define that a total mapping $\delta' : S \to S$ leads to a v-loop if and only if $\delta'(s)$ has a v-loop for all $s \in S$. We may also say for short that $w \in A^*$ leads to a v-loop if δ^w leads to a v-loop. Moreover, for $v, w \in A^*$ we write $v \sim_F w$ if and only if $\delta^v = \delta^w$.

Recall the following theorem concerning starfree languages.

Theorem 2.1 [Sch65, MP71]. *Let $F = (A, S, \delta, s_0, S')$ be a minimal dfa. Then $L(F)$ is starfree if and only if there is some $m \geq 0$ such that for all $w \in A^+$ and for all $p \in S$ it holds that $\delta(p, w^m) = \delta(p, w^{m+1})$.*

We call a minimal dfa F permutationfree if it has the above property. The decision of this property for given F is known to be PSPACE–complete [CH91]. For later use we restate the previous theorem as follows.

Proposition 2.2. *Let $F = (A, S, \delta, s_0, S')$ be a minimal dfa. $L(F)$ is not starfree if and only if there exist $w \in A^+$, some $l \geq 2$ and distinct states $r_0, r_1, \ldots, r_{l-1} \in S$ such that $\delta(r_i, w) = r_{i+1}$ for $0 \leq i \leq l-1$ (with $r_l =_{\mathrm{def}} r_0$).*

An obvious property of permutationfree dfa's is that they run into a w-loop after a small number of successive w's in the input.

Proposition 2.3. *Let $F = (A, S, \delta, s_0, S')$ be a permutationfree dfa and $r \geq |F|$. Then $\delta(s, w^r) = \delta(s, w^{r+1})$ for all $w \in A^*$ and $s \in S$.*

Proof. Since $\{ \delta(s, w^i) \mid 0 \leq i \leq r+1 \} \subseteq S$ there must be $0 \leq i < j \leq r+1$ with $s' =_{\mathrm{def}} \delta(s, w^i) = \delta(s, w^j)$. Then $s' = \delta(s, w^l)$ for $i \leq l \leq j$ since otherwise F is not permutationfree. It follows that $s' = \delta(s, w^l)$ for $i \leq l \leq r+1$ and in particular $\delta(s, w^r) = \delta(s, w^{r+1})$ because $i \leq r$. ❑

Proposition 2.4. *Let $w \in A^*$ and $r \geq 1$, then w^r leads to a $w^{r!}$-loop in every dfa F with $|F| \leq r$.*

Proof. Observe that w^r leads to a w^i-loop for some $1 \leq i \leq |F|$. The proposition follows since every such w^i-loop can be considered as a $w^{r!}$-loop. ❑

2.2 Definition of Concatenation Hierarchies

We study in this paper two well–known concatenation hierarchies, namely the dot–depth hierarchy (DDH, for short) and the Straubing–Thérien hierarchy (STH, for short). For a class $\mathcal{C}$ of languages we denote its closure under finite (possibly empty) union by $\mathrm{FU}(\mathcal{C})$. Set

$$\mathrm{Pol}(\mathcal{C}) =_{\mathrm{def}} \mathrm{FU}(\{L_0 L_1 \cdots L_n : n \geq 0 \text{ and } L_i \in \mathcal{C}\})$$

as the polynomial closure of $\mathcal{C}$. Note that $\mathrm{Pol}(\mathcal{C})$ is exactly the closure of $\mathcal{C}$ under finite (possibly empty) union and finite (non–empty) concatenation. Furthermore, $\mathcal{C}$ is a subset of the polynomial closure of $\mathcal{C}$. For a second closure operation we consider Boolean operations. We take A^+ as our universe and denote the Boolean closure of a class $\mathcal{C}$ of languages of A^+ by $\mathrm{BC}(\mathcal{C})$ (taking complements with respect to A^+).

Definition 2.5 (DDH). *We define the levels of the dot–depth hierarchy as follows:*

$$\begin{array}{lll} \mathcal{B}_{1/2} & =_{\mathrm{def}} & \mathrm{Pol}(\{\{w\} : w \in A^+\} \cup \{A^+\}) & \\ \mathcal{B}_{n+1} & =_{\mathrm{def}} & \mathrm{BC}(\mathcal{B}_{n+1/2}) & \text{for } n \geq 0 \\ \mathcal{B}_{n+3/2} & =_{\mathrm{def}} & \mathrm{Pol}(\mathcal{B}_{n+1}) & \text{for } n \geq 0 \end{array}$$

Definition 2.6 (STH). *We define the levels of the Straubing–Thérien hierarchy as follows:*

$$\begin{array}{lll} \mathcal{L}_{1/2} & =_{\mathrm{def}} & \mathrm{Pol}(\{A^* a A^* : a \in A\}) & \\ \mathcal{L}_{n+1} & =_{\mathrm{def}} & \mathrm{BC}(\mathcal{L}_{n+1/2}) & \text{for } n \geq 0 \\ \mathcal{L}_{n+3/2} & =_{\mathrm{def}} & \mathrm{Pol}(\mathcal{L}_{n+1}) & \text{for } n \geq 0 \end{array}$$

The discussion in the forthcoming Section 2.3 relates our definitions to the ones known from literature. The following inclusion relations in each hierarchy are easy to see from the definitions.

Proposition 2.7. *For $n \geq 0$ the following holds.*

1. $\mathcal{B}_{n+1/2} \cup \mathrm{co}\mathcal{B}_{n+1/2} \subseteq \mathcal{B}_{n+1} \subseteq \mathcal{B}_{n+3/2} \cap \mathrm{co}\mathcal{B}_{n+3/2}$
2. $\mathcal{L}_{n+1/2} \cup \mathrm{co}\mathcal{L}_{n+1/2} \subseteq \mathcal{L}_{n+1} \subseteq \mathcal{L}_{n+3/2} \cap \mathrm{co}\mathcal{L}_{n+3/2}$

We can also compare the hierarchies in both directions by inclusion.

Proposition 2.8. *For $n \geq 1$ the following holds.*

1. $\mathcal{L}_{n-1/2} \subseteq \mathcal{B}_{n-1/2} \subseteq \mathcal{L}_{n+1/2}$
2. $\mathrm{co}\mathcal{L}_{n-1/2} \subseteq \mathrm{co}\mathcal{B}_{n-1/2} \subseteq \mathrm{co}\mathcal{L}_{n+1/2}$
3. $\mathcal{L}_n \subseteq \mathcal{B}_n \subseteq \mathcal{L}_{n+1}$

Proof. Since for all $a \in A$ it holds that $A^* a A^* = \{a\} \cup aA \cup Aa \cup A^+ a A^+ \in \mathcal{B}_{1/2}$ we obtain $\mathcal{L}_{1/2} \subseteq \mathcal{B}_{1/2}$. Moreover, it holds that $A^+ = \bigcup_{a \in A} A^* a A^* \in \mathcal{L}_{1/2}$, and for $w \in A^+$ with $w = a_1 \cdots a_n$ for letters $a_i \in A$ and $n \geq 1$ we obtain

$$\{w\} = \underbrace{A^* a_1 A^* \cdots a_n A^*}_{\in \mathcal{L}_{1/2}} \cap \left(A^+ \setminus \underbrace{\bigcup_{b_1, \ldots, b_{n+1} \in A} A^* b_1 A^* \cdots b_{n+1} A^*}_{\in \mathcal{L}_{1/2}} \right) \in \mathcal{L}_1.$$

In particular $A^+ \in \mathcal{L}_{3/2}$ and $w \in \mathcal{L}_{3/2}$ for all $w \in A^+$ from which we get $\mathcal{B}_{1/2} \subseteq \mathcal{L}_{3/2}$. So we have seen $\mathcal{L}_{1/2} \subseteq \mathcal{B}_{1/2} \subseteq \mathcal{L}_{3/2}$ and the proposition follows from the monotony of Pol and BC, and complementation. ❑

It is shown in [Eil76] that $\bigcup_{n \geq 1} \mathcal{B}_n = SF$. Together with Proposition 2.8 we get the following.

Proposition 2.9. $\bigcup_{n \geq 1} \mathcal{L}_{n/2} = \bigcup_{n \geq 1} \mathcal{B}_{n/2} = SF$.

2.3 Alternative Definitions

The dot–depth hierarchy and the Straubing–Thérien hierarchy have gained much attention due to the still pending dot–depth problem. The purpose of this section is to make our work comparable to other investigations, so we discuss alternative definitions. There are two points to look at: First, one finds other versions of the polynomial closure operation in the literature. Let $\mathcal{C}$ be a class of languages. Here are the definitions of polynomial closure as chosen, e.g., with an algebraic approach recently in [PW97].

$$\begin{aligned}\mathrm{Pol}^{\mathcal{L}}(\mathcal{C}) &=_{\mathrm{def}} \mathrm{FU}(\{L_0a_1L_1\cdots a_nL_n : n\geq 0, L_i\in\mathcal{C} \text{ and } a_i\in A\})\\ \mathrm{Pol}^{\mathcal{B}}(\mathcal{C}) &=_{\mathrm{def}} \mathrm{FU}(\{u_0L_1u_1\cdots L_nu_n : n\geq 0, L_i\in\mathcal{C}, u_i\in A^* \text{ and if } n=0 \text{ then } u_0\neq\varepsilon\})\end{aligned}$$

A second point is that languages may be defined in a way such that they contain the empty word. So we also want to see if that makes any difference. It is pointed out, e.g., in [Pin95] that this is a crucial point in the theory of varieties of finite semigroups. Since most results in the field were obtained via this theory, the following definitions of concatenation hierarchies are widely used. We denote the Boolean closure of a class $\mathcal{D}$ of languages of A^* by $\mathrm{BC}^*(\mathcal{D})$ (taking complements with respect to A^*).

Definition 2.10 (DDH due to [Pin96]). *Let $\mathcal{B}^+_{1/2}$ be the class of all languages of A^+ which can be written as finite unions of languages of the form $u_0A^+u_1\cdots A^+u_m$ where $m\geq 0$ and $u_i\in A^*$. For $n\geq 0$ let $\mathcal{B}^+_{n+1} =_{\mathrm{def}} \mathrm{BC}(\mathcal{B}^+_{n+1/2})$ and $\mathcal{B}^+_{n+3/2} =_{\mathrm{def}} \mathrm{Pol}^{\mathcal{B}}(\mathcal{B}^+_{n+1})$.*

Definition 2.11 (STH due to [Str81, Thé81]). *Let $\mathcal{L}^*_{1/2}$ be the class of all languages of A^* which can be written as finite unions of languages of the form $A^*a_1A^*\cdots a_mA^*$ where $m\geq 0$ and $a_i\in A$. For $n\geq 0$ let $\mathcal{L}^*_{n+1} =_{\mathrm{def}} \mathrm{BC}^*(\mathcal{L}^*_{n+1/2})$ and $\mathcal{L}^*_{n+3/2} =_{\mathrm{def}} \mathrm{Pol}^{\mathcal{L}}(\mathcal{L}^*_{n+1})$.*

These definitions are local to the remainder of Section 2. In this paper, we follow an automata–theoretic approach, and we find it suitable to have the inclusion relations from Proposition 2.8 at hand. Theorem 2.20 below shows that the classes $\mathcal{B}^+_{n/2}$ from Definition 2.10 and our classes $\mathcal{B}_{n/2}$ coincide, and that the languages in $\mathcal{L}^*_{n/2}$ are up to the empty word the languages in $\mathcal{L}_{n/2}$. Let us recall known closure properties of the just defined classes.

Lemma 2.12 [Arf91, PW97, Gla98]. *Let $n\geq 1$ and $a\in A$.*

1. *The classes $\mathcal{B}^+_{n/2}, \mathrm{co}\mathcal{B}^+_{n/2}, \mathcal{L}^*_{n/2}$ and $\mathrm{co}^*\mathcal{L}^*_{n/2}$ are closed under finite union and intersection.*
2. *Let $\mathcal{C}$ be one of the classes $\mathcal{B}^+_{n/2}$ or $\mathrm{co}\mathcal{B}^+_{n/2}$. Then $a^{-1}L\cap A^+, La^{-1}\cap A^+\in\mathcal{C}$ for $L\in\mathcal{C}$.*
3. *Let $\mathcal{C}$ be one of the classes $\mathcal{L}^*_{n/2}$ or $\mathrm{co}^*\mathcal{L}^*_{n/2}$. Then $a^{-1}L, La^{-1}\in\mathcal{C}$ for $L\in\mathcal{C}$.*

2.3.1 Polynomial Closure

We investigate the relationships between the different notions of polynomial closure, and identify a condition for $\mathcal{C}$ under which these notions coincide.

Theorem 2.13. $\mathrm{Pol}(\mathcal{C}) = \mathrm{Pol}^{\mathcal{L}}(\mathcal{C}) = \mathrm{Pol}^{\mathcal{B}}(\mathcal{C})$ *for a class of languages $\mathcal{C}$ satisfying the conditions:*

(a) $\{w\}\in\mathcal{C}$ *for every word* $w\in A^+$

(b) $a^{-1}L\cap A^+\in\mathcal{C}$ *and* $La^{-1}\cap A^+\in\mathcal{C}$ *for every* $L\in\mathcal{C}$ *and* $a\in A$

The proof is an easy consequence of the following lemma.

Lemma 2.14. *Let $\mathcal{C}$ be a class of languages. Then the following holds.*

1. $\mathrm{Pol}^{\mathcal{L}}(\mathcal{C})\subseteq\mathrm{Pol}^{\mathcal{B}}(\mathcal{C})$

2. $\mathrm{Pol}(\mathcal{C}) \subseteq \mathrm{Pol}^{\mathcal{B}}(\mathcal{C})$

3. *If* $\{w\} \in \mathcal{C}$ *for all* $w \in A^+$, *then* $\mathrm{Pol}^{\mathcal{B}}(\mathcal{C}) \subseteq \mathrm{Pol}(\mathcal{C})$.

4. *If* $\{w\}, a^{-1}L \cap A^+, La^{-1} \cap A^+ \in \mathcal{C}$ *for all* $w \in A^1 \cup A^2$, $a \in A$, $L \in \mathcal{C}$, *then* $\mathrm{Pol}(\mathcal{C}) \subseteq \mathrm{Pol}^{\mathcal{L}}(\mathcal{C})$.

5. *If* $\{a\} \in \mathcal{C}$ *for all* $a \in A$ *then* $\mathrm{Pol}^{\mathcal{L}}(\mathcal{C}) \subseteq \mathrm{Pol}(\mathcal{C})$.

Proof. The statements 1, 2 and 5 can be easily verified. If we are given a language of the form $u_0L_1u_1 \cdots L_nu_n$ with $n \geq 0$, languages $L_i \in \mathcal{C}$ and words $u_i \in A^*$ such that $n = 0 \Longrightarrow u_0 \neq \varepsilon$, then we can first take out every $u_i = \varepsilon$ from this representation without changing the language. If $\{w\} \in \mathcal{C}$ for all $w \in A^+$, then we can replace all remaining $u_i \in A^+$ by languages from $\mathcal{C}$. We obtain an equivalent expression of the form $L'_0L'_1 \cdots L'_{n'}$ with $n' \geq 0$ and $L'_i \in \mathcal{C}$ (note that if $n = 0$, then $\{u_0\} \in \mathcal{C}$). This shows statement 3.

Let us turn to statement 4. Here we have $\{w\} \in \mathcal{C}$ for all words $w \in A^+$ with length 1 or 2, and $a^{-1}L \cap A^+, La^{-1} \cap A^+ \in \mathcal{C}$ for letters $a \in A$ and languages $L \in \mathcal{C}$. It suffices to show that $L_0L_1 \cdots L_n \in \mathrm{Pol}^{\mathcal{L}}(\mathcal{C})$ for $n \geq 0$ and $L_i \in \mathcal{C}$. We will prove this by induction on n. For $n = 0$ (induction base) this is trivial. We assume that we have proven statement 4 for $n = m$ and we want to show that $L_0L_1 \cdots L_{m+1} \in \mathrm{Pol}^{\mathcal{L}}(\mathcal{C})$ for $m \geq 0$ and $L_i \in \mathcal{C}$. Since $L_0L_1 \cdots L_m \in \mathrm{Pol}^{\mathcal{L}}(\mathcal{C})$ by induction hypothesis, it suffices to show that $L =_{\mathrm{def}} L'_0a_1L'_1 \cdots a_lL'_l \cdot L_{m+1} \in \mathrm{Pol}^{\mathcal{L}}(\mathcal{C})$ for $l \geq 0$, $L'_i \in \mathcal{C}$ and $a_i \in A$. With $L' =_{\mathrm{def}} L'_0a_1L'_1 \cdots a_lL'_l$ we obtain

$$L = \begin{cases} \left(\bigcup\limits_{a \in A} L' \cdot a \cdot \left(a^{-1}L_{m+1} \cap A^+\right)\right) \cup \left(\bigcup\limits_{a \in L_{m+1} \cap A} L' \cdot a\right) \cup L' & : \text{ if } \varepsilon \in L_{m+1} \\ \left(\bigcup\limits_{a \in A} L' \cdot a \cdot \left(a^{-1}L_{m+1} \cap A^+\right)\right) \cup \left(\bigcup\limits_{a \in L_{m+1} \cap A} L' \cdot a\right) & : \text{ otherwise.} \end{cases}$$

By assumption we have $a^{-1}L_{m+1} \cap A^+ \in \mathcal{C}$. It follows that

$$\bigcup_{a \in A} L' \cdot a \cdot \left(a^{-1}L_{m+1} \cap A^+\right) \in \mathrm{Pol}^{\mathcal{L}}(\mathcal{C}).$$

Since also $L' \in \mathrm{Pol}^{\mathcal{L}}(\mathcal{C})$, it remains to show that $L' \cdot b \in \mathrm{Pol}^{\mathcal{L}}(\mathcal{C})$ for letters $b \in A$. If $l > 0$ then define $L'' =_{\mathrm{def}} L'_0a_1 \cdots L'_{l-1}a_l$. Now consider the following case study of $L' \cdot b$.

$$L' \cdot b = \begin{cases} \left(\bigcup\limits_{a \in A} L'' \cdot \left(L'_la^{-1} \cap A^+\right) \cdot a \cdot b\right) \cup \left(\bigcup\limits_{a \in L'_l \cap A} L'' \cdot a \cdot b\right) \cup L'' \cdot b & : \text{ if } l > 0 \text{ and } \varepsilon \in L'_l \\ \left(\bigcup\limits_{a \in A} L'' \cdot \left(L'_la^{-1} \cap A^+\right) \cdot a \cdot b\right) \cup \left(\bigcup\limits_{a \in L'_l \cap A} L'' \cdot a \cdot b\right) & : \text{ if } l > 0 \text{ and } \varepsilon \notin L'_l \\ \left(\bigcup\limits_{a \in A} \left(L'_la^{-1} \cap A^+\right) \cdot a \cdot b\right) \cup \left(\bigcup\limits_{a \in L'_l \cap A} \{ab\}\right) \cup \{b\} & : \text{ if } l = 0 \text{ and } \varepsilon \in L'_l \\ \left(\bigcup\limits_{a \in A} \left(L'_la^{-1} \cap A^+\right) \cdot a \cdot b\right) \cup \left(\bigcup\limits_{a \in L'_l \cap A} \{ab\}\right) & : \text{ if } l = 0 \text{ and } \varepsilon \notin L'_l \end{cases}$$

By assumption, $\{b\}, \{ab\} \in \mathcal{C}$ for letters a, b and $\tilde{L} =_{\mathrm{def}} L'_la^{-1} \cap A^+ \in \mathcal{C}$. Hence for $a, b \in A$ the following holds.

- $L'' \cdot \left(L'_la^{-1} \cap A^+\right) \cdot a \cdot b = L'_0a_1 \cdots L'_{l-1}a_l \cdot \tilde{L} \cdot a \cdot \{b\} \in \mathrm{Pol}^{\mathcal{L}}(\mathcal{C})$ for $l > 0$
- $L'' \cdot a \cdot b = L'_0a_1 \cdots L'_{l-1}a_l \cdot \{ab\} \in \mathrm{Pol}^{\mathcal{L}}(\mathcal{C})$ for $l > 0$
- $L'' \cdot b = L'_0a_1 \cdots L'_{l-1}a_l \cdot \{b\} \in \mathrm{Pol}^{\mathcal{L}}(\mathcal{C})$ for $l > 0$

- $\left(L'_l a^{-1} \cap A^+\right) \cdot a \cdot b = \tilde{L} \cdot a \cdot \{b\} \in \mathrm{Pol}^{\mathcal{L}}(\mathcal{C})$

Together with the case study this implies $L' \cdot b \in \mathrm{Pol}^{\mathcal{L}}(\mathcal{C})$. This completes both the induction and the proof of statement 4. ❑

If the languages of two classes $\mathcal{C}$ and $\mathcal{D}$ differ only by the empty word, we show that this property is preserved by the polynomial closure operation. In other words, we obtain that also $\mathrm{Pol}(\mathcal{C})$ is equal to $\mathrm{Pol}(\mathcal{D})$ up to the empty word.

Lemma 2.15. *Let $\mathcal{C}$ be a class of languages of A^+ and $\mathcal{D}$ be a class of languages of A^*. If $\{\varepsilon\} \in \mathcal{D}$ and $\mathcal{D} = \mathcal{C} \cup \{L \cup \{\varepsilon\} : L \in \mathcal{C}\}$ then the following holds.*

1. $\mathrm{Pol}(\mathcal{C})$ *is a class of languages of* A^+ *and* $\mathrm{Pol}(\mathcal{D})$ *is a class of languages of* A^*
2. $\{\varepsilon\} \in \mathrm{Pol}(\mathcal{D})$
3. $\mathrm{Pol}(\mathcal{D}) = \mathrm{Pol}(\mathcal{C}) \cup \{L \cup \{\varepsilon\} : L \in \mathrm{Pol}(\mathcal{C})\}$

Proof. The statements 1 and 2 follow immediately from the definition of Pol. Now we want to show $\mathrm{Pol}(\mathcal{D}) \subseteq \mathrm{Pol}(\mathcal{C}) \cup \{L \cup \{\varepsilon\} : L \in \mathrm{Pol}(\mathcal{C})\}$. Since the right hand side is closed under finite union, it suffices to show the following claim.
Claim: $L_0 \cdots L_n \in \mathrm{Pol}(\mathcal{C}) \cup \{L \cup \{\varepsilon\} : L \in \mathrm{Pol}(\mathcal{C})\}$ for $n \geq 0$ and $L_i \in \mathcal{D}$.

We will show this by induction on n. Observe that for $n = 0$ (induction base) the claim holds. Now assume that we have already proven the claim for the case $n = m$ and we want to show it for $n = m+1$. Let $L =_{\mathrm{def}} L_0 \cdots L_{m+1}$ with $L_i \in \mathcal{D}$. By induction hypothesis there exists an $L' \in \mathrm{Pol}(\mathcal{C})$, such that $L_0 \cdots L_m = L'$ or $L_0 \cdots L_m = L' \cup \{\varepsilon\}$. Since $L_{m+1} \in \mathcal{D}$, there exists an $L'' \in \mathcal{C} \subseteq \mathrm{Pol}(\mathcal{C})$ such that $L_{m+1} = L''$ or $L_{m+1} = L'' \cup \{\varepsilon\}$. It follows that

$$L = \begin{cases} L'L'' & : \text{ if } L_0 \cdots L_m = L' \text{ and } L_{m+1} = L'' \\ L'L'' \cup L' & : \text{ if } L_0 \cdots L_m = L' \text{ and } L_{m+1} = L'' \cup \{\varepsilon\} \\ L'L'' \cup L'' & : \text{ if } L_0 \cdots L_m = L' \cup \{\varepsilon\} \text{ and } L_{m+1} = L'' \\ L'L'' \cup L'' \cup L' \cup \{\varepsilon\} & : \text{ if } L_0 \cdots L_m = L' \cup \{\varepsilon\} \text{ and } L_{m+1} = L'' \cup \{\varepsilon\}\,. \end{cases}$$

This shows that in any case there exists an $\tilde{L} \in \mathrm{Pol}(\mathcal{C})$ such that $L = \tilde{L}$ or $L = \tilde{L} \cup \{\varepsilon\}$. This proves the claim.

Finally we have to show the inclusion $\mathrm{Pol}(\mathcal{D}) \supseteq \mathrm{Pol}(\mathcal{C}) \cup \{L \cup \{\varepsilon\} : L \in \mathrm{Pol}(\mathcal{C})\}$. Since $\mathcal{D} \supseteq \mathcal{C}$, we have $\mathrm{Pol}(\mathcal{D}) \supseteq \mathrm{Pol}(\mathcal{C})$. Furthermore, we have $L \cup \{\varepsilon\} \in \mathrm{Pol}(\mathcal{D})$ for $L \in \mathrm{Pol}(\mathcal{C})$, since $\mathrm{Pol}(\mathcal{C}) \subseteq \mathrm{Pol}(\mathcal{D})$ and $\{\varepsilon\} \in \mathrm{Pol}(\mathcal{D})$. This proves the lemma. ❑

2.3.2 Boolean Closure

We show a counterpart of Lemma 2.15, this time for Boolean closures.

Lemma 2.16. *Let $\mathcal{C}$ be a class of languages of A^+ and $\mathcal{D}$ be a class of languages of A^*. If $\{\varepsilon\} \in \mathcal{D}$ and $\mathcal{D} = \mathcal{C} \cup \{L \cup \{\varepsilon\} : L \in \mathcal{C}\}$ then the following holds.*

1. $\mathrm{BC}(\mathcal{C})$ *is a class of languages of* A^+ *and* $\mathrm{BC}^*(\mathcal{D})$ *is a class of languages of* A^*
2. $\{\varepsilon\} \in \mathrm{BC}^*(\mathcal{D})$
3. $\mathrm{BC}^*(\mathcal{D}) = \mathrm{BC}(\mathcal{C}) \cup \{L \cup \{\varepsilon\} : L \in \mathrm{BC}(\mathcal{C})\}$

Proof. The statements 1 and 2 follow from the definition of BC and BC*. Now we want to show the inclusion $\mathrm{BC}^*(\mathcal{D}) \subseteq \mathrm{BC}(\mathcal{C}) \cup \{L \cup \{\varepsilon\} : L \in \mathrm{BC}(\mathcal{C})\}$. Note that elements from $\mathrm{BC}^*(\mathcal{D})$ can be written as finite unions of finite intersections of literals which in turn are of the form L (positive literal) or $A^* \setminus L$ (negative literal) for some $L \in \mathcal{D}$. Observe that the right hand side of the equation in statement 3 is closed under finite union and intersection. Thus it suffices to show that all literals are elements of the right hand side. Since $\mathcal{D} \subseteq \mathcal{C} \cup \{L \cup \{\varepsilon\} : L \in \mathcal{C}\} \subseteq \mathrm{BC}(\mathcal{C}) \cup \{L \cup \{\varepsilon\} : L \in \mathrm{BC}(\mathcal{C})\}$, all positive literals are elements of the right hand side, and it remains to show the following claim.
Claim: $A^* \setminus L \in \mathrm{BC}(\mathcal{C}) \cup \{L \cup \{\varepsilon\} : L \in \mathrm{BC}(\mathcal{C})\}$ for all $L \in \mathcal{D}$.

Let $L \in \mathcal{D}$ then there exists an $L' \in \mathcal{C}$ such that $L = L'$ or $L = L' \cup \{\varepsilon\}$. Since $L' \subseteq A^+$, we obtain $A^* \setminus L = (A^+ \setminus L') \cup \{\varepsilon\}$ in the first case, and $A^* \setminus L = (A^+ \setminus L')$ in the second case. Since $A^+ \setminus L' \in \mathrm{BC}(\mathcal{C})$, we conclude that $A^* \setminus L \in \mathrm{BC}(\mathcal{C}) \cup \{L \cup \{\varepsilon\} : L \in \mathrm{BC}(\mathcal{C})\}$. This proves our claim, and it follows that $\mathrm{BC}^*(\mathcal{D}) \subseteq \mathrm{BC}(\mathcal{C}) \cup \{L \cup \{\varepsilon\} : L \in \mathrm{BC}(\mathcal{C})\}$.

Let us turn to the reverse inclusion. From $\mathcal{C} \subseteq \mathcal{D}$, $\{\varepsilon\} \in \mathrm{BC}^*(\mathcal{D})$ and $A^+ \setminus L = (A^* \setminus L) \cap (A^* \setminus \{\varepsilon\})$ for $L \subseteq A^+$, it follows that $\mathrm{BC}^*(\mathcal{D}) \supseteq \mathrm{BC}(\mathcal{C})$ and $\mathrm{BC}^*(\mathcal{D}) \supseteq \{L \cup \{\varepsilon\} : L \in \mathrm{BC}(\mathcal{C})\}$. This shows the inclusion $\mathrm{BC}^*(\mathcal{D}) \supseteq \mathrm{BC}(\mathcal{C}) \cup \{L \cup \{\varepsilon\} : L \in \mathrm{BC}(\mathcal{C})\}$. ❑

2.3.3 Comparing the Lower Levels

To apply Lemma 2.15 and 2.16 for the general case, we need to compare the hierarchies on their lower levels. This will be the induction base for the proof of Theorem 2.20 below.

Proposition 2.17. *The following holds.*

1. $\mathcal{B}_{1/2} = \mathcal{B}^+_{1/2}$
2. $\mathcal{B}_{1/2}$ *equals the class of languages of* A^+ *which can be written as finite unions of languages of the form* $u_0A^*u_1 \cdots A^*u_m$ *where* $m \geq 0$ *and* $u_i \in A^*$.

Proof. Observe that elements of $\mathcal{B}_{1/2}$ are languages of A^+ which can be written as finite unions of languages of the form $u_0A^+u_1 \cdots A^+u_m$ where $m \geq 0$ and $u_i \in A^*$ (add some $u_i = \varepsilon$ if necessary). It follows that $\mathcal{B}_{1/2} \subseteq \mathcal{B}^+_{1/2}$. On the other hand, languages of A^+ having the form $u_0A^+u_1 \cdots A^+u_m$ with $m \geq 0$ and $u_i \in A^*$, can be written as concatenations of A^+ and non–empty words (just drop all $u_i = \varepsilon$). This shows $\mathcal{B}_{1/2} \supseteq \mathcal{B}^+_{1/2}$ and statement 1 follows.

Now we can exploit statement 1 for the proof of statement 2. Therefore, it suffices to show that for each language $L \subseteq A^+$ the following holds: L is a finite union of languages of the form $u_0A^*u_1 \cdots A^*u_m$ with $m \geq 0$, $u_i \in A^*$ if and only if L is a finite union of languages of the form $u_0A^+u_1 \cdots A^+u_m$ with $m \geq 0$, $u_i \in A^*$. This is easy to see since we can replace A^* and A^+ due to $A^* = \{\varepsilon\} \cup A^+$ and $A^+ = \bigcup_{a \in A} aA^*$. This shows statement 2. ❑

Proposition 2.18. *The following holds.*

1. $\mathcal{L}_{1/2}$ *equals the class of languages of* A^+ *which can be written as finite unions of languages of the form* $A^*a_1A^* \cdots a_mA^*$ *where* $m \geq 0$ *and* $a_i \in A$.
2. $\mathcal{L}_{1/2}$ *is a class of languages of* A^+ *and* $\mathcal{L}^*_{1/2} = \mathcal{L}_{1/2} \cup \{A^*\}$.
3. $\mathcal{L}_1$ *is a class of languages of* A^+ *and* $\mathcal{L}^*_1 = \mathcal{L}_1 \cup \{L \cup \{\varepsilon\} : L \in \mathcal{L}_1\}$.

Proof. The first parts of all statements are by definition. To see the remaining parts of statement 1 it suffices to mention that (i) the case $m = 0$ can not occur since we speak about languages of A^+ and (ii) $A^* = A^*A^*$. The second part of statement 2 follows immediately from statement 1 and the definition of $\mathcal{L}^*_{1/2}$.

For the second part of statement 3, observe that $\mathcal{L}_1^*$ and $\mathcal{L}_1 \cup \{L \cup \{\varepsilon\} : L \in \mathcal{L}_1\}$ are classes being closed under finite union and finite intersection. Furthermore, we have

$$\{\varepsilon\} = \bigcap_{a \in A} A^* \setminus A^*aA^* \in \mathcal{L}_1^*.$$

Thus it suffices to show that (i) $L, A^* \setminus L \in \mathcal{L}_1 \cup \{L \cup \{\varepsilon\} : L \in \mathcal{L}_1\}$ for $L \in \mathcal{L}_{1/2}^*$ and (ii) $L', A^+ \setminus L' \in \mathcal{L}_1^*$ for $L' \in \mathcal{L}_{1/2}$. Since by the second statement of the lemma we have $\mathcal{L}_{1/2}^* = \mathcal{L}_{1/2} \cup \{A^*\}$, and since

$$A^* = \underbrace{\bigcup_{a \in A} A^*aA^*}_{\in \mathcal{L}_{1/2}} \cup \{\varepsilon\},$$

it follows that $\mathcal{L}_{1/2}^* \subseteq \mathcal{L}_1 \cup \{L \cup \{\varepsilon\} : L \in \mathcal{L}_1\}$ and $\mathcal{L}_{1/2} \subseteq \mathcal{L}_1^*$. Hence it remains to show the following.

1. $A^* \setminus L \in \mathcal{L}_1 \cup \{L \cup \{\varepsilon\} : L \in \mathcal{L}_1\}$ for $L \in \mathcal{L}_{1/2}^*$
2. $A^+ \setminus L' \in \mathcal{L}_1^*$ for $L' \in \mathcal{L}_{1/2}$

First of all let $L \in \mathcal{L}_{1/2}^*$. If $L \notin \mathcal{L}_{1/2}$ then $L = A^*$ and we obtain $A^* \setminus L \in \mathcal{L}_{1/2} \subseteq \mathcal{L}_1$. Otherwise we can write $A^* \setminus L = (A^+ \setminus L) \cup \{\varepsilon\} \in \{L \cup \{\varepsilon\} : L \in \mathcal{L}_1\}$.

Now let $L' \in \mathcal{L}_{1/2} \subseteq \mathcal{L}_{1/2}^*$. Since $A^+ \in \mathcal{L}_{1/2}^*$ and $A^+ \setminus L' = (A^* \setminus L') \cap A^+$, we obtain $A^+ \setminus L' \in \mathcal{L}_1^*$. This proves the proposition. ❑

2.3.4 Comparing the Hierarchies

We apply our Theorem 2.13.

Proposition 2.19.

1. *For $n \geq 1$, let $\mathcal{C}$ be one of the classes $\mathcal{B}_{n/2}^+$ or $\mathrm{co}\mathcal{B}_{n/2}^+$. Then* $\mathrm{Pol}(\mathcal{C}) = \mathrm{Pol}^{\mathcal{L}}(\mathcal{C}) = \mathrm{Pol}^{\mathcal{B}}(\mathcal{C})$.
2. *For $n \geq 2$, let $\mathcal{C}$ be one of the classes $\mathcal{L}_{n/2}^*$ or $\mathrm{co}\mathcal{L}_{n/2}^*$. Then* $\mathrm{Pol}(\mathcal{C}) = \mathrm{Pol}^{\mathcal{L}}(\mathcal{C}) = \mathrm{Pol}^{\mathcal{B}}(\mathcal{C})$.

Proof. To show this we need to prove that the mentioned classes fulfil the condition from Theorem 2.13.

(a) $\{w\} \in \mathcal{C}$ for every word $w \in A^+$

(b) $a^{-1}L \cap A^+ \in \mathcal{C}$ and $La^{-1} \cap A^+ \in \mathcal{C}$ for every $L \in \mathcal{C}$ and $a \in A$

So let $w \in A^+$. Then we have $\{w\} \in \mathcal{B}_{1/2}^+$ by definition, and we see that

$$\{w\} = A^+ \setminus \left(\underbrace{\bigcup_{\substack{v \in A^+ \text{with} \\ |v| = |w|}} vA^+\varepsilon \cup \bigcup_{\substack{v \in A^+ \setminus \{w\} \\ \text{with } |v| \leq |w|}} v}_{\in \mathcal{B}_{1/2}^+} \right) \in \mathrm{co}\mathcal{B}_{1/2}^+.$$

This shows (a) for statement 1 due to the known inclusions. If $w = a_1 \cdots a_n$ for $n \geq 1$ and letters $a_i \in A$, we obtain (a) for statement 2 with

$$\{w\} = \underbrace{A^*a_1A^* \cdots a_nA^*}_{\in \mathcal{L}_{1/2}^*} \cap \left(A^* \setminus \underbrace{\bigcup_{b_1, \ldots, b_{n+1} \in A} A^*b_1A^* \cdots b_{n+1}A^*}_{\in \mathcal{L}_{1/2}^*} \right) \in \mathcal{L}_1^*.$$

Condition (b) is fulfilled for statement 1 by Lemma 2.12. We use the same lemma for statement 2 together with the closure under intersection. So it suffices to note that $A^+ \in \mathcal{L}_{1/2}^* \subseteq \mathcal{L}_1^*$. ❑

Theorem 2.20. *The following holds.*

1. $\mathcal{B}^+_{n/2} = \mathcal{B}_{n/2}$ *for* $n \geq 1$
2. $\mathcal{L}^*_{1/2} = \mathcal{L}_{1/2} \cup \{A^*\}$
3. $\mathcal{L}^*_{n/2} = \mathcal{L}_{n/2} \cup \big\{L \cup \{\varepsilon\} : L \in \mathcal{L}_{n/2}\big\}$ *for* $n \geq 2$

Proof. We show statement 1 by induction on n. By Proposition 2.17, the assertion holds for $n = 1$ (induction base). We first assume that it holds for $n = m \geq 1$ with $m \equiv 1 \mod 2$, and we want to prove it for $m + 1$. Then

$$\begin{aligned} \mathcal{B}_{(m+1)/2} &= \mathrm{BC}(\mathcal{B}_{m/2}) \\ &= \mathrm{BC}(\mathcal{B}^+_{m/2}) \qquad (1) \\ &= \mathcal{B}^+_{(m+1)/2}, \end{aligned}$$

where we use the induction hypothesis $\mathcal{B}_{m/2} = \mathcal{B}^+_{m/2}$ to obtain equation (1). This shows in particular statement 1 for $n = 2$. So now we assume that it holds for $n = m \geq 2$ with $m \equiv 0 \mod 2$, and we want to prove it for $m + 1$. Then we have $\mathcal{B}_{(m+1)/2} = \mathrm{Pol}(\mathcal{B}_{m/2})$, and from Proposition 2.19 we get $\mathcal{B}^+_{(m+1)/2} = \mathrm{Pol}^{\mathcal{B}}(\mathcal{B}^+_{m/2}) = \mathrm{Pol}(\mathcal{B}^+_{m/2})$. The induction hypothesis provides $\mathrm{Pol}(\mathcal{B}_{m/2}) = \mathrm{Pol}(\mathcal{B}^+_{m/2})$.

Statement 2 is given in Proposition 2.18.

We prove statement 3 also by induction on n. For $n = 2$, the induction base is given in Proposition 2.18. We first assume that it holds for $n = m \geq 2$ with $m \equiv 0 \mod 2$, and we want to prove it for $m + 1$. Then we have $\mathcal{L}^*_{(m+1)/2} = \mathrm{Pol}^{\mathcal{L}}(\mathcal{L}^*_{m/2})$ and $\mathcal{L}_{(m+1)/2} = \mathrm{Pol}(\mathcal{L}_{m/2})$. It holds that $\mathcal{L}_{m/2}$ is a class of languages of A^+, $\{\varepsilon\} \in \mathcal{L}^*_1 \subseteq \mathcal{L}^*_{m/2}$, and from the induction hypothesis we obtain $\mathcal{L}^*_{m/2} = \mathcal{L}_{m/2} \cup \big\{L \cup \{\varepsilon\} : L \in \mathcal{L}_{m/2}\big\}$. This shows that the classes $\mathcal{C} =_{\mathrm{def}} \mathcal{L}_{m/2}$ and $\mathcal{D} =_{\mathrm{def}} \mathcal{L}^*_{m/2}$ satisfy the assumptions of Lemma 2.15 and we obtain

$$\mathrm{Pol}(\mathcal{L}^*_{m/2}) = \mathrm{Pol}(\mathcal{L}_{m/2}) \cup \big\{L \cup \{\varepsilon\} : L \in \mathrm{Pol}(\mathcal{L}_{m/2})\big\}. \qquad (2)$$

Because $m \geq 2$ we get from Proposition 2.19 that $\mathrm{Pol}(\mathcal{L}^*_{m/2}) = \mathrm{Pol}^{\mathcal{L}}(\mathcal{L}^*_{m/2}) = \mathcal{L}^*_{(m+1)/2}$. This shows in particular statement 3 for $n = 3$. So now we assume that it holds for $n = m \geq 3$ with $m \equiv 1 \mod 2$, and we want to prove it for $m + 1$. Then we have $\mathcal{L}^*_{(m+1)/2} = \mathrm{BC}^*(\mathcal{L}^*_{m/2})$ and $\mathcal{L}_{(m+1)/2} = \mathrm{BC}(\mathcal{L}_{m/2})$. By definition, $\mathcal{L}_{m/2}$ is a class of languages of A^+, and still $\{\varepsilon\} \in \mathcal{L}^*_1 \subseteq \mathcal{L}^*_{m/2}$. Furthermore, from the induction hypothesis we obtain $\mathcal{L}^*_{m/2} = \mathcal{L}_{m/2} \cup \big\{L \cup \{\varepsilon\} : L \in \mathcal{L}_{m/2}\big\}$. Thus the classes $\mathcal{C} =_{\mathrm{def}} \mathcal{L}_{m/2}$ and $\mathcal{D} =_{\mathrm{def}} \mathcal{L}^*_{m/2}$ satisfy the assumptions of Lemma 2.16 and we obtain

$$\mathrm{BC}^*(\mathcal{L}^*_{m/2}) = \mathrm{BC}(\mathcal{L}_{m/2}) \cup \big\{L \cup \{\varepsilon\} : L \in \mathrm{BC}(\mathcal{L}_{m/2})\big\}.$$

This shows

$$\mathcal{L}^*_{(m+1)/2} = \mathcal{L}_{(m+1)/2} \cup \big\{L \cup \{\varepsilon\} : L \in \mathcal{L}_{(m+1)/2}\big\}.$$

❑

Let us carry over Theorem 2.20 to the classes of complements.

Corollary 2.21. *The following holds.*

1. $\mathrm{co}\mathcal{B}^+_{n+1/2} = \mathrm{co}\mathcal{B}_{n+1/2}$ *for* $n \geq 0$
2. $\mathrm{co}^*\mathcal{L}^*_{n+1/2} = \mathrm{co}\mathcal{L}_{n+1/2} \cup \big\{L \cup \{\varepsilon\} : L \in \mathrm{co}\mathcal{L}_{n+1/2}\big\}$ *for* $n \geq 1$

Proof. Statement 1 is an immediate consequence of Theorem 2.20, from which we also get for $n \geq 1$:

$$\begin{aligned} \mathrm{co}^*\mathcal{L}^*_{n+1/2} = \left\{A^* \setminus L : L \in \mathcal{L}^*_{n+1/2}\right\} &= \overbrace{\left\{A^* \setminus L : L \in \mathcal{L}_{n+1/2}\right\}}^{\text{languages } L \text{ with } \varepsilon \in L} \cup \overbrace{\left\{A^+ \setminus L : L \in \mathcal{L}_{n+1/2}\right\}}^{\text{languages } L \text{ with } \varepsilon \notin L} \\ &= \left\{L \cup \{\varepsilon\} : L \in \mathrm{co}\mathcal{L}_{n+1/2}\right\} \cup \mathrm{co}\mathcal{L}_{n+1/2} \end{aligned}$$

This shows the second statement. ❑

Note that $\mathcal{L}_{1/2}$ and $\mathcal{L}^*_{1/2}$ in Theorem 2.20 are some kind of exception since all classes $\mathcal{L}^*_{n/2}$ with $n \geq 2$ have the following property: $L \cup \{\varepsilon\} \in \mathcal{L}^*_{n/2} \iff L \setminus \{\varepsilon\} \in \mathcal{L}^*_{n/2}$ for all languages $L \subseteq A^*$. This does not hold for $\mathcal{L}^*_{1/2}$ because A^* is the only language in $\mathcal{L}^*_{1/2}$ which contains the empty word. However, we have the following uniform statement of this relation.

Corollary 2.22. *For $n \geq 1$ we have $\mathcal{L}_{n/2} = \mathcal{L}^*_{n/2} \cap \mathcal{P}(A^+)$.*

Proof. For $n \geq 2$ this follows from Theorem 2.20. By definition, $\mathcal{L}_{1/2}$ is a class of languages of A^+, and if we intersect both sides of $\mathcal{L}^*_{1/2} = \mathcal{L}_{1/2} \cup \{A^*\}$ with $\mathcal{P}(A^+)$ we get $\mathcal{L}_{1/2} = \mathcal{L}^*_{1/2} \cap \mathcal{P}(A^+)$. ❑

2.4 Normalforms and Closure Properties

Finally, we give in this preliminary section some normalforms and closure properties for the hierarchy classes $\mathcal{L}_{n/2}$ and $\mathcal{B}_{n/2}$. We first mention the normalform for $\mathcal{L}_{3/2}$ from [Arf87].

Proposition 2.23.

1. *$\mathcal{L}_{3/2}$ is equal to the class of languages of A^+ which can be written as finite unions of languages of the form $A_0^* a_1 A_0^* \cdots a_m A_m^*$ where $m \geq 0$, $a_i \in A$ and $A_i \subseteq A$.*
2. *$\mathcal{L}_{3/2}$ is equal to the class of languages of A^+ which can be written as finite unions of languages of the form $u_0 A_1^+ u_1 \cdots A_m^+ u_m$ where $m \geq 0$, $u_i \in A^*$ and $\emptyset \neq A_i \subseteq A$.*

Proof. In [Arf87] it is shown that $\mathcal{L}^*_{3/2}$ is equal to the class of languages of A^* which can be written as finite unions of languages of the form $A_0^* a_1 A_0^* \cdots a_m A_m^*$ where $m \geq 0$, $a_i \in A$ and $A_i \subseteq A$. By Corollary 2.22 we have

$$\mathcal{L}_{3/2} = \left\{L \in \mathcal{L}^*_{3/2} : L \subseteq A^+\right\}$$

which shows statement 1. Observe that $\emptyset^* = \{\varepsilon\}$, $A'^* = A'^+ \cup \{\varepsilon\}$ and $A'^+ = \bigcup_{a \in A'} aA'^*$ for $a \in A$ and $\emptyset \neq A' \subseteq A$. So it is easy to see by mutual substitution that the following statements are equivalent for every language $L \subseteq A^+$.

- L is a finite union of languages of the form $A_0^* a_1 A_0^* \cdots a_m A_m^*$ with $m \geq 0$, $a_i \in A$ and $A_i \subseteq A$.
- L is a finite union of languages of the form $u_0 A_1^* u_1 \cdots A_m^* u_m$ with $m \geq 0$, $u_i \in A^*$, $\emptyset \neq A_i \subseteq A$.
- L is a finite union of languages of the form $u_0 A_1^+ u_1 \cdots A_m^+ u_m$ with $m \geq 0$, $u_i \in A^*$, $\emptyset \neq A_i \subseteq A$.

This shows statement 2. ❑

In [Gla98] normalforms for the definition of levels $n + 1/2$ of the dot–depth hierarchy and the Straubing–Thérien hierarchy are proven.

Lemma 2.24. *For $n \geq 1$ the following holds.*

1. $\mathcal{L}_{n+1/2} = \mathrm{Pol}(\mathrm{co}\mathcal{L}_{n-1/2})$
2. $\mathcal{B}_{n+1/2} = \mathrm{Pol}(\mathrm{co}\mathcal{B}_{n-1/2})$

Proof. By definition, $\mathcal{L}_n \supseteq \mathrm{co}\mathcal{L}_{n-1/2}$ and $\mathcal{B}_n \supseteq \mathrm{co}\mathcal{B}_{n-1/2}$ for $n \geq 1$. Thus we have $\mathcal{L}_{n+1/2} \supseteq \mathrm{Pol}(\mathrm{co}\mathcal{L}_{n-1/2})$ and $\mathcal{B}_{n+1/2} \supseteq \mathrm{Pol}(\mathrm{co}\mathcal{B}_{n-1/2})$ for $n \geq 1$. It remains to show the other inclusions.

For this end, we recall from [Gla98] that for $n \geq 1$ it holds that

$$\begin{aligned} \mathcal{L}^*_{n+1/2} &= \mathrm{Pol}^{\mathcal{L}}(\mathrm{co}^*\mathcal{L}^*_{n-1/2}) && (3)\\ \mathcal{B}^+_{n+1/2} &= \mathrm{Pol}^{\mathcal{B}}(\mathrm{co}\mathcal{B}^+_{n-1/2}) && (4) \end{aligned}$$

First of all we consider statement 1 for $n = 1$. By Proposition 2.23, languages from $\mathcal{L}_{3/2}$ can be written as finite unions of languages of the form $u_0 A_1^+ u_1 \cdots A_m^+ u_m$ where $m \geq 0$, $u_i \in A^*$ and $\emptyset \neq A_i \subseteq A$. Note that if $m = 0$ then $u_0 \neq \varepsilon$, since languages of $\mathcal{L}_{3/2}$ do not contain the empty word. Hence it suffices to show that $A'^+, \{a\} \in \mathrm{co}\mathcal{L}_{1/2}$ for $\emptyset \neq A' \subseteq A$ and all $a \in A$. This can be seen as follows.

$$\begin{aligned} A'^+ &= A^+ \setminus \bigcup_{a \in A \setminus A'} A^* a A^* \in \mathrm{co}\mathcal{L}_{1/2} \\ \{a\} &= A^+ \setminus \Bigg(\underbrace{\bigcup_{a' \in A \setminus \{a\}} A^* a' A^*}_{\substack{\text{words of length } \geq 1 \\ \text{containing a letter } a' \neq a}} \cup \underbrace{A^* a A^* a A^*}_{\text{words with } \geq 2\ a'\text{s}} \Bigg) \in \mathrm{co}\mathcal{L}_{1/2} \end{aligned}$$

This shows $\mathcal{L}_{3/2} \subseteq \mathrm{Pol}(\mathrm{co}\mathcal{L}_{1/2})$. Now we consider statement 1 for some $n \geq 2$. Here we have $\mathrm{co}^*\mathcal{L}^*_{n-1/2} = \mathrm{co}\mathcal{L}_{n-1/2} \cup \{L \cup \{\varepsilon\} : L \in \mathrm{co}\mathcal{L}_{n-1/2}\}$ by Theorem 2.20. Since $\{\varepsilon\} \in \mathrm{co}^*\mathcal{L}^*_{1/2} \subseteq \mathrm{co}^*\mathcal{L}^*_{n-1/2}$, we can apply Lemma 2.15 as before and we obtain

$$\mathrm{Pol}(\mathrm{co}^*\mathcal{L}^*_{n-1/2}) = \mathrm{Pol}(\mathrm{co}\mathcal{L}_{n-1/2}) \cup \{L \cup \{\varepsilon\} : L \in \mathrm{Pol}(\mathrm{co}\mathcal{L}_{n-1/2})\}. \tag{5}$$

From Proposition 2.19 we see that $\mathrm{Pol}(\mathrm{co}^*\mathcal{L}^*_{n-1/2}) = \mathrm{Pol}^{\mathcal{L}}(\mathrm{co}^*\mathcal{L}^*_{n-1/2})$. So together with (3) we can rewrite (5) as

$$\mathcal{L}^*_{n+1/2} = \underbrace{\mathrm{Pol}(\mathrm{co}\mathcal{L}_{n-1/2})}_{\text{languages } L \text{ with } \varepsilon \notin L} \cup \underbrace{\{L \cup \{\varepsilon\} : L \in \mathrm{Pol}(\mathrm{co}\mathcal{L}_{n-1/2})\}}_{\text{languages } L \text{ with } \varepsilon \in L}. \tag{6}$$

We can compare this to Theorem 2.20 where we have

$$\mathcal{L}^*_{n+1/2} = \underbrace{\mathcal{L}_{n+1/2}}_{\text{languages } L \text{ with } \varepsilon \notin L} \cup \underbrace{\{L \cup \{\varepsilon\} : L \in \mathcal{L}_{n+1/2}\}}_{\text{languages } L \text{ with } \varepsilon \in L}. \tag{7}$$

Because the unions in (6) and (7) are disjoint, we see that $\mathcal{L}_{n+1/2} = \mathrm{Pol}(\mathrm{co}\mathcal{L}_{n-1/2})$ and statement 1 is proven.

Let us consider statement 2 for $n \geq 1$. From (4) and Theorem 2.20 we obtain $\mathcal{B}_{n+1/2} = \mathrm{Pol}^{\mathcal{B}}(\mathrm{co}\mathcal{B}^+_{n-1/2})$. Together with Proposition 2.19 this yields $\mathcal{B}_{n+1/2} = \mathrm{Pol}(\mathrm{co}\mathcal{B}^+_{n-1/2})$. With Corollary 2.21 we get $\mathcal{B}_{n+1/2} = \mathrm{Pol}(\mathrm{co}\mathcal{B}_{n-1/2})$. ❑

Now we translate the closure properties from Lemma 2.12 to our definitions.

Lemma 2.25. *Let $n \geq 1$ and $a \in A$.*

1. *The classes $\mathcal{B}_{n/2}, \mathrm{co}\mathcal{B}_{n/2}, \mathcal{L}_{n/2}$ and $\mathrm{co}\mathcal{L}_{n/2}$ are closed under finite union and intersection.*
2. *Let $\mathcal{C}$ be one of the classes $\mathcal{B}_{n/2}, \mathrm{co}\mathcal{B}_{n/2}, \mathcal{L}_{n/2}$ or $\mathrm{co}\mathcal{L}_{n/2}$. Then $a^{-1}L \cap A^+, La^{-1} \cap A^+ \in \mathcal{C}$ for $L \in \mathcal{C}$.*

Proof. For the classes $\mathcal{B}_{n/2}$ and $\mathrm{co}\mathcal{B}_{n/2}$ the lemma follows from Theorem 2.20 and Lemma 2.12.

The closure of $\mathcal{L}_{n/2}$ under finite union and intersection for $n \geq 1$ is immediate from Lemma 2.12 and Corollary 2.22. This carries over to $\mathrm{co}\mathcal{L}_{n/2}$.

Now let $n \geq 1$, $a \in A$ and $L \in \mathcal{L}_{n/2}$. By Theorem 2.20 we have $\mathcal{L}_{n/2} \subseteq \mathcal{L}^*_{n/2}$. Thus $L \in \mathcal{L}^*_{n/2}$ and we obtain $La^{-1}, a^{-1}L \in \mathcal{L}^*_{n/2}$ by Lemma 2.12. Since $A^+ \in \mathcal{L}_{1/2} \subseteq \mathcal{L}_{n/2}$ it follows from the closure under intersection and from Theorem 2.20 that $La^{-1} \cap A^+, a^{-1}L \cap A^+ \in \mathcal{L}_{n/2}$. Analogously this can be shown for $n \geq 1$, $a \in A$ and $L \in \mathrm{co}\mathcal{L}_{n+1/2}$ using Corollary 2.21 and Lemma 2.12.

Finally let $L' \in \mathrm{co}\mathcal{L}_{1/2}$ with $L' = A^+ \setminus L$ for some $L \in \mathcal{L}_{1/2}$. Then we have

$$\begin{aligned} L'a^{-1} \cap A^+ = \{v \in A^+ : va \in L'\} &= \{v \in A^+ : va \in A^+ \setminus L\} \\ &= \{v \in A^+ : va \notin L\} = A^+ \setminus \{v \in A^+ : va \in L\} \\ &= A^+ \setminus \underbrace{(La^{-1} \cap A^+)}_{\in \mathcal{L}_{1/2}} \in \mathrm{co}\mathcal{L}_{1/2}. \end{aligned}$$

Analogously one shows $a^{-1}L \cap A^+ \in \mathrm{co}\mathcal{L}_{1/2}$. ❑

3 A Theory of Forbidden Patterns

The dot–depth hierarchy and the Straubing–Thérien hierarchy are hierarchies of classes of regular languages, and it is known from literature that their lower levels have characterizations in terms of forbidden patterns. In this section we take up this idea and develop a method for a uniform definition of hierarchies via iterated forbidden patterns. Such a definition starts with a so–called initial pattern which determines the first level of the corresponding hierarchy. Using an iteration rule we obtain more complicated patterns which in turn define the higher levels.

In subsection 3.1 we introduce the notion of iterated patterns, and in subsection 3.2 we prove some auxiliary results which help to handle these patterns in a better way. In subsection 3.3 we prove an interesting relation between the polynomial closure operation on the one hand and our iteration rule on the other hand. More precisely we show that a complementation followed by a polynomial closure operation on the language side is captured by our iteration rule on the forbidden pattern side. In subsection 3.4 we investigate the general inclusion structure between hierarchies which are defined via iterated forbidden patterns. In subsection 3.5 we study in how far the pattern iteration rule preserves starfreeness, and in subsection 3.6 we treat the decidability of our pattern hierarchies.

3.1 Definition of Iterated Patterns

Starting with a so–called initial pattern (fulfilling some reasonable weak prerequisites) and using a certain iteration rule, we define hierarchies of patterns in dfa's. In later subsections we will use these hierarchies of patterns for the definition of hierarchies of languages which share some interesting properties with concatenation hierarchies.

Definition 3.1. *We define an initial pattern $\mathcal{T}$ to be a subset of $A^* \times A^*$ such that for all $r \geq 1$ and $v, w \in A^*$ it holds that $(v, w) \in \mathcal{T} \Longrightarrow (v, v), (v^r, w \cdot v^r) \in \mathcal{T}$.*

Definition 3.2. *For every set $\mathbb{L}$ we define $\mathrm{IT}(\mathbb{L}) =_{\mathrm{def}} \{(w_0, p_0, \ldots, w_m, p_m) : p_i \in \mathbb{L}, w_i \in A^+\}$.*

Definition 3.3. *For an initial pattern $\mathcal{T}$ we define $\mathbb{L}_0^{\mathcal{T}} =_{\mathrm{def}} \mathcal{T}$ and $\mathbb{L}_{i+1}^{\mathcal{T}} =_{\mathrm{def}} \mathrm{IT}(\mathbb{L}_i^{\mathcal{T}})$ for $i \geq 0$.*

Definition 3.4. *Let $\mathcal{T}$ be an initial pattern, $F = (A, S, \delta, s_0, S')$ a dfa and $s, s_1, s_2 \in S$.*

1. *For $p = (v, w) \in \mathbb{L}_0^{\mathcal{T}}$ we define:*
   - *(a) $\overline{p} =_{\mathrm{def}} w$ and $\overline{p}^{\circ} =_{\mathrm{def}} v$*
   - *(b) p appears at state $s \overset{\mathrm{def}}{\Longleftrightarrow} \delta(s, v) = s$*
   - *(c) the states s_1, s_2 are connected via $p \overset{\mathrm{def}}{\Longleftrightarrow}$ p appears at s_1 and s_2, and $\delta(s_1, w) = s_2$*
2. *For $n \geq 0$ and $p = (w_0, p_0, \ldots, w_m, p_m) \in \mathbb{L}_{n+1}^{\mathcal{T}}$ we define the following:*
   - *(a) $\overline{p} =_{\mathrm{def}} w_0 \cdots w_m$ and $\overline{p}^{\circ} =_{\mathrm{def}} w_0\overline{p_0} \cdots w_m\overline{p_m}$*
   - *(b) p appears at state $s \overset{\mathrm{def}}{\Longleftrightarrow}$ there exist states $\tilde{r}_i, \tilde{s}_i \in S$ with $0 \leq i \leq m$ such that $\tilde{r}_0 = \delta(s, w_0)$, $\tilde{s}_m = s$, $\delta(\tilde{s}_j, w_{j+1}) = \tilde{r}_{j+1}$ for $0 \leq j < m$ and the states $\tilde{r}_i, \tilde{s}_i$ are connected via p_i for $0 \leq i \leq m$.*
   - *(c) the states s_1, s_2 are connected via $p \overset{\mathrm{def}}{\Longleftrightarrow}$ p appears at state s_1, p appears at state s_2 and there exist states $\tilde{s}_i \in S$ with $0 \leq i \leq m$ such that $\tilde{s}_0 = \delta(s_1, w_0)$, $\tilde{s}_m = s_2$, $\delta(\tilde{s}_j, w_{j+1}) = \tilde{s}_{j+1}$ for $0 \leq j < m$ and p_i appears at state $\tilde{s}_i$ for $0 \leq i \leq m$*

We give the following interpretation of the definitions made so far. In the known forbidden patterns for concatenation hierarchies there usually appear two states s_1, s_2 both having a loop with the same structure in the dfa (loop–structure for short). This structure in turn determines the subgraph we need to find between s_1 and s_2 (bridge–structure). Each pattern class $\mathbb{L}_n^{\mathcal{T}}$ with $n \geq 0$ can be understood as a set of structures in dfa's, and each such structure determines a certain loop–structure and a certain bridge–structure. If a $p \in \mathbb{L}_n^{\mathcal{T}}$ appears at some state s, then this means that we can find the loop–structure described by p at s. If two states s_1, s_2 are connected via p, then (i) p appears at both states and (ii) we find the bridge–structure described by p between them. To handle these patterns, we define for $p \in \mathbb{L}_n^{\mathcal{T}}$ a word $\overline{p}^\circ$ obtained from the loop–structure described by p (call this the loop–word), and a word $\overline{p}$ which is derived from the bridge–structure of p (bridge–word). It will turn out that if p appears at some state s, then this state has a $\overline{p}^\circ$-loop. If s_1 and s_2 are connected via p then the bridge–word $\overline{p}$ leads directly from s_1 to s_2.

Let us make this more precise, first for an initial pattern. Let $p = (v, w) \in \mathcal{T}$ be an element of an initial pattern $\mathcal{T}$. Then we interpret v as a loop (the loop–structure of p) and w as a certain path (the bridge–structure of p) in a dfa. Therefore, the loop–word $\overline{p}^\circ$ is defined as v and the bridge–word $\overline{p}$ is defined as w. So p appears at state s if and only if this state has a v-loop. Two states s_1, s_2 are connected via p if and only if (i) both states have a v-loop and (ii) w leads from s_1 to s_2, i.e., $\delta(s_1, w) = s_2$.

Now let p be an element of $\mathbb{L}_n^{\mathcal{T}}$ for $n > 0$. This means that $p = (w_0, p_0, \ldots, w_m, p_m)$ for words w_i and elements $p_i \in \mathbb{L}_{n-1}^{\mathcal{T}}$. We interpret the loop–structure described by p as a loop which can be divided in words $w_0, w_1, \ldots, w_m$ in this ordering, and between each w_i, w_{i+1} we find p_i (i.e., the states at the end of each w_i and at the beginning of each w_{i+1} are connected via p_i). Here we see that elements of $\mathbb{L}_{n-1}^{\mathcal{T}}$ appear in the loop–structure of elements of $\mathbb{L}_n^{\mathcal{T}}$. This is a phenomenon which can also be observed in the known forbidden pattern characterizations for the levels $1/2$ and $3/2$ of the dot–depth hierarchy and the Straubing–Thérien hierarchy (cf. Figures 4, 5, 7, 8). The bridge–structure of p can be interpreted as a path in a dfa labeled with $w_0 w_1 \cdots w_m$ and going from state s_1 to s_2 such that (i) we find the loop–structure of p at both states and (ii) after the word $w_0 \cdots w_i$ we reach a state where we find the loop–structure of p_i.

The known forbidden pattern characterizations for levels $1/2$ and $3/2$ of concatenation hierarchies describe patterns which consist of reachable states s_1, s_2 having certain structural properties such that some word z is accepted when starting in s_1 and rejected when starting in s_2. The following definition takes this into account.

Definition 3.5. *Let $F = (A, S, \delta, s_0, S')$ be a dfa, $\mathcal{T}$ an initial pattern and $n \geq 0$. We say that F has pattern $\mathbb{L}_n^{\mathcal{T}}$ if and only if there exist $s_1, s_2 \in S$, $u, z \in A^*$, $p \in \mathbb{L}_n^{\mathcal{T}}$ such that $\delta(s_0, u) = s_1$, $\delta(s_1, z) \in S'$, $\delta(s_2, z) \notin S'$ and the states s_1, s_2 are connected via p.*

3.2 Auxiliary Results for Iterated Patterns

In order to establish a relation between the polynomial closure operation and the iteration rule for the patterns classes (cf. Theorem 3.15), we isolate the main argument of this proof in Lemma 3.11 below, for which the following two constructions are needed. First, for $p \in \mathbb{L}_n^{\mathcal{T}}$ some $\lambda(p) \in \mathbb{L}_n^{\mathcal{T}}$ can be defined such that if p appears at some state s then s, s are connected via $\lambda(p)$ (cf. Definition 3.7 and Lemma 3.8). Secondly, in Definition 3.9 and Lemma 3.10 we pump up the loop–structure of p to construct for given $r \geq 1$ some $\pi(p, r) \in \mathbb{L}_n^{\mathcal{T}}$ such that

(i) if two states are connected via p, then they are also connected via $\pi(p, r)$ and

(ii) in every dfa F' with $|F'| \leq r$ the words $\overline{\pi(p, r)}$ and $\overline{\pi(p, r)}^\circ$ lead to states where $\pi(p, r)$ appears.

First of all let us states some basic properties of iterated patterns.

Lemma 3.6. *Let $\mathcal{T}$ be an initial pattern, $n \geq 0$, $p \in \mathbb{L}_n^{\mathcal{T}}$, $F = (A, S, \delta, s_0, S')$ a dfa and $s, s_1, s_2 \in S$.*

1. *If s_1, s_2 are connected via p then p appears at s_1 and at s_2.*

2. *If s_1, s_2 are connected via p then $\delta(s_1, \overline{p}) = s_2$.*

3. *If p appears at state s then $\delta(s, \overline{p}^\circ) = s$.*

4. *If $n > 0$ then $\overline{p}, \overline{p}^\circ \in A^+$.*

5. *Let $n > 0$ and $p = (w_0, p_0, \ldots, w_m, p_m)$ for suitable $m \geq 0$, $w_i \in A^+$ and $p_i \in \mathbb{L}^{\mathcal{T}}_{n-1}$. If p appears at state s then also p_m appears at state s.*

Proof. We only have to consider statement 3, since the statements 1, 2, 4 and 5 are easy consequences from the definition. Let $p \in \mathbb{L}^{\mathcal{T}}_n$ for some $n \geq 0$ such that p appears at state s. If $n = 0$ then we have $\delta(s, \overline{p}^\circ) = \delta(s, v) = s$. If $n > 0$ then there exist states $\tilde{r}_i, \tilde{s}_i \in S$ with $0 \leq i \leq m$ such that $\tilde{r}_0 = \delta(s, w_0)$, $\tilde{s}_m = s$, $\delta(\tilde{s}_j, w_{j+1}) = \tilde{r}_{j+1}$ for $0 \leq j < m$ and the states $\tilde{r}_i, \tilde{s}_i$ are connected via p_i for $0 \leq i \leq m$. From statement 2 it follows that $\delta(\tilde{r}_i, \overline{p_i}) = \tilde{s}_i$ for $0 \leq i \leq m$. This implies:

$$\begin{aligned}
\delta(s, \overline{p}^\circ) = \delta(s, w_0\overline{p_0}w_1\overline{p_1}\cdots w_m\overline{p_m}) &= \delta(\tilde{r}_0, \overline{p_0}w_1\overline{p_1}\cdots w_m\overline{p_m}) \\
&= \delta(\tilde{s}_0, w_1\overline{p_1}\cdots w_m\overline{p_m}) \\
&= \delta(\tilde{r}_1, \overline{p_1}\cdots w_m\overline{p_m}) \\
&\vdots \\
&= \delta(\tilde{s}_{m-1}, w_m\overline{p_m}) = \delta(\tilde{r}_m, \overline{p_m}) = \tilde{s}_m = s
\end{aligned}$$

❑

Let $\mathcal{T}$ be an initial pattern. Given some $r \geq 3$, a dfa F and states $s_1, s_2 \in F$ which are connected via some $p \in \mathbb{L}^{\mathcal{T}}_n$, we want to construct a $p' \in \mathbb{L}^{\mathcal{T}}_n$ such that (i) the occurrence of p' is implied by the occurrence of p and (ii) $\overline{p'}$ and $\overline{p'}^\circ$ lead to a state where p' appears in every dfa F' with $|F'| \leq r$. First of all we will see that a $p \in \mathbb{L}^{\mathcal{T}}_n$ appearing at state s can be considered as a connection between s and s. For this we have to carry out a slight modification of p. For $n \geq 0$ and $p \in \mathbb{L}^{\mathcal{T}}_n$ we will construct a similar pattern $\lambda(p) \in \mathbb{L}^{\mathcal{T}}_n$ such that the following holds for every state s of some dfa F: If p appears at s, then s, s are connected via $\lambda(p)$.

Definition 3.7. *Let $\mathcal{T}$ be an initial pattern. For $p = (v, w) \in \mathbb{L}^{\mathcal{T}}_0$ we define $\lambda(p) =_{\text{def}} (v, v)$. Now let $p \in \mathbb{L}^{\mathcal{T}}_{n+1}$ such that for some $m \geq 0$, $p_i \in \mathbb{L}^{\mathcal{T}}_n$ and $w_i \in A^+$ we have $p = (w_0, p_0, \ldots, w_m, p_m)$. We define $\lambda(p) =_{\text{def}} (\overline{p}^\circ, \lambda(p_m))$.*

Lemma 3.8. *Let $\mathcal{T}$ be an initial pattern, $n \geq 0$ and $p \in \mathbb{L}^{\mathcal{T}}_n$, then $\lambda(p) \in \mathbb{L}^{\mathcal{T}}_n$. Furthermore, let $F = (A, S, \delta, s_0, S')$ be a dfa and $s \in S$ such that p appears at state s. Then the states s, s are connected via $\lambda(p)$.*

Proof. We prove the lemma by induction on n. For $n = 0$ we have $p = (v, w) \in \mathbb{L}^{\mathcal{T}}_0 = \mathcal{T}$. It follows from Definition 3.1 that $\lambda(p) = (v, v) \in \mathcal{T} = \mathbb{L}^{\mathcal{T}}_0$. Since p appears at state s, we have $\delta(s, v) = s$. Therefore, the states s, s are connected via $\lambda(p) = (v, v)$ by Definition 3.4.

We assume that the lemma has been shown for all $n \leq l$, we want to prove it for $n = l + 1$. Let $p \in \mathbb{L}^{\mathcal{T}}_{l+1}$ such that for some $m \geq 0$, $p_i \in \mathbb{L}^{\mathcal{T}}_l$ and $w_i \in A^+$ we have $p = (w_0, p_0, \ldots, w_m, p_m)$. By Lemma 3.6.4 we have $\overline{p}^\circ \in A^+$, and from induction hypothesis we know that $\lambda(p_m) \in \mathbb{L}^{\mathcal{T}}_l$. From the Definitions 3.2 and 3.3 it follows that $\lambda(p) = (\overline{p}^\circ, \lambda(p_m)) \in \mathrm{IT}(\mathbb{L}^{\mathcal{T}}_l) = \mathbb{L}^{\mathcal{T}}_{l+1}$.

It remains to show that the states s, s are connected via $\lambda(p) = (\overline{p}^\circ, \lambda(p_m))$ in dfa F. By Lemma 3.6.5, p_m appears at state s. From the induction hypothesis it follows that s, s are connected via $\lambda(p_m)$. Since $\delta(s, \overline{p}^\circ) = s$ (Lemma 3.6.3), we obtain that $\lambda(p)$ appears at state s. Now let $s_1 =_{\text{def}} s$, $s_2 =_{\text{def}} s$ and $\tilde{s}_0 =_{\text{def}} s$. Since p appears at state s, it follows from Lemma 3.6.3 that $\tilde{s}_0 = \delta(s_1, \overline{p}^\circ)$. We have already seen that s, s are connected via $\lambda(p_m)$, particularly $\lambda(p_m)$ appears at state $s = \tilde{s}_0$ (Lemma 3.6.1). This shows that s, s are connected via $\lambda(p)$. ❑

Definition 3.9. *Let $\mathcal{T}$ be an initial pattern. For $r \geq 3$ and $p = (v, w) \in \mathbb{L}_0^{\mathcal{T}}$ we define $\pi(p, r) =_{\rm def} (v^{r!}, w \cdot v^{r!})$. Now let $n \geq 0$ and $p \in \mathbb{L}_{n+1}^{\mathcal{T}}$ such that for some $m \geq 0$, $p_i \in \mathbb{L}_n^{\mathcal{T}}$ and $w_i \in A^+$ we have $p = (w_0, p_0, \ldots, w_m, p_m)$. For $r \geq 3$ we define the following.*

$$\begin{aligned} p_i' &=_{\rm def} \pi(p_i, r) \\ w &=_{\rm def} w_0 \cdot \overline{p_0'}^{\circ} \cdot \overline{p_0'} \cdots w_m \cdot \overline{p_m'}^{\circ} \cdot \overline{p_m'} \\ \pi(p, r) &=_{\rm def} (w_0 \cdot \overline{p_0'}^{\circ}, p_0', \ldots, w_m \cdot \overline{p_m'}^{\circ}, p_m', \underbrace{w, \lambda(p_m'), \ldots, w, \lambda(p_m')}_{(r!-1) \text{ times } \text{“}w, \lambda(p_m')\text{”}}) \end{aligned}$$

Lemma 3.10. *Let $\mathcal{T}$ be an initial pattern, $n \geq 0$, $r \geq 3$, $p \in \mathbb{L}_n^{\mathcal{T}}$ and $p' =_{\rm def} \pi(p, r)$.*

1. *$p' \in \mathbb{L}_n^{\mathcal{T}}$*
2. *If p appears at state s of some dfa F, then also p' appears at this state.*
3. *If the states s_1, s_2 of some dfa F are connected via p, then these states are also connected via p'.*
4. *If F' is a dfa with $|F'| \leq r$, then $\overline{p'}^{\circ}$ and $\overline{p'}$ lead to a state in F' where p' appears.*
5. *If F' is a dfa with $|F'| \leq r$, then $\overline{p'}^{\circ}, \overline{p'}^{\circ}\overline{p'}$ lead to states in F' which are connected via p'.*
6. *If F' is a dfa with $|F'| \leq r$, then $\overline{p'}, \overline{p'}\,\overline{p'}$ lead to states in F' which are connected via p'.*

Proof. Let $\mathcal{T}$ be an initial pattern and $r \geq 3$. We will prove the lemma by induction on n.
Induction Base:
For $n = 0$ we have $p = (v, w) \in \mathbb{L}_0^{\mathcal{T}} = \mathcal{T}$ and $p' = (v^{r!}, w \cdot v^{r!})$. From Definition 3.1 it follows that $p' \in \mathcal{T} = \mathbb{L}_0^{\mathcal{T}}$. Let $F = (A, S, \delta, s_0, S')$ be dfa and $s, s_1, s_2 \in S$. If p appears at state s, then we have $\delta(s, v) = s$. Hence $\delta(s, v^{r!}) = s$ and it follows that p' appears at state s. If s_1, s_2 are connected via p, then $s_1 = \delta(s_1, v)$ and $s_2 = \delta(s_1, w) = \delta(s_2, v)$. It follows that $s_1 = \delta(s_1, v^{r!})$ and $s_2 = \delta(s_1, w \cdot v^{r!}) = \delta(s_2, v^{r!})$. Thus s_1, s_2 are also connected via p'. By Proposition 2.4, $v^{r!} = v^{r!-r} \cdot v^r$ leads to a $v^{r!}$-loop in every dfa F' with $|F'| \leq r$. Since $\overline{p'}^{\circ} = v^{r!}$ and $\overline{p'} = w \cdot v^{r!}$, in every dfa F' with $|F'| \leq r$ both words $\overline{p'}^{\circ}$ and $\overline{p'}$ lead to states which have a $v^{r!}$-loop. It follows that p' appears at these states. Furthermore, it is easy to see that in every dfa F' with $|F'| \leq r$ the words $\overline{p'}^{\circ}, \overline{p'}^{\circ}\overline{p'}$ lead to states s_1', s_2' (resp.), such that $\delta(s_1', v^{r!}) = s_1'$ and $\delta(s_1', w \cdot v^{r!}) = \delta(s_2', v^{r!}) = s_2'$. Hence s_1', s_2' are connected via p'. Analogously we see that $\overline{p'}, \overline{p'}\,\overline{p'}$ lead to states which are connected via p'. This shows the induction base.
Induction Step:
Suppose now that we have proven the lemma for $n = l$ and we want to show it for $n = l + 1$. Let $p \in \mathbb{L}_{l+1}^{\mathcal{T}}$, $p' =_{\rm def} \pi(p, r)$ and choose suitable $m \geq 0$, $p_i \in \mathbb{L}_l^{\mathcal{T}}$, $w_i \in A^+$ such that $p = (w_0, p_0, \ldots, w_m, p_m)$. As in Definition 3.9 let $p_i' =_{\rm def} \pi(p_i, r)$ and $w =_{\rm def} w_0 \cdot \overline{p_0'}^{\circ} \cdot \overline{p_0'} \cdots w_m \cdot \overline{p_m'}^{\circ} \cdot \overline{p_m'}$. Moreover let $F = (A, S, \delta, s_0, S')$ be a dfa and $s, s_1, s_2 \in S$ such that (i) p appears at state s and (ii) the states s_1, s_2 are connected via p.

First of all let us prove that $\delta(s', w) = s'$ for every state $s' \in S$ where p appears. If p appears at s', then there exist states $\tilde{r}_i, \tilde{s}_i \in S$ with $0 \leq i \leq m$ such that $\tilde{r}_0 = \delta(s', w_0)$, $\tilde{s}_m = s'$, $\delta(\tilde{s}_j, w_{j+1}) = \tilde{r}_{j+1}$ for $0 \leq j < m$ and the states $\tilde{r}_i, \tilde{s}_i$ are connected via p_i for $0 \leq i \leq m$. By induction hypothesis, $\tilde{r}_i, \tilde{s}_i$ are also connected via p_i' for $0 \leq i \leq m$. From Lemma 3.6.2 it follows that $\delta(\tilde{r}_i, \overline{p_i'}) = \tilde{s}_i$ for $0 \leq i \leq m$. Hence $\delta(s', w_0 \cdot \overline{p_0'} \cdots w_{i-1} \cdot \overline{p_{i-1}'} \cdot w_i) = \tilde{r}_i$ and $\delta(s', w_0 \cdot \overline{p_0'} \cdots w_{i-1} \cdot \overline{p_{i-1}'} \cdot w_i \cdot \overline{p_i'}) = \tilde{s}_i$ for $0 \leq i \leq m$. By the Lemmas 3.6.1 and 3.6.3, the state $\tilde{r}_i$ has a $\overline{p_i'}^{\circ}$-loop for $0 \leq i \leq m$. It follows that $\delta(s', w_0 \cdot \overline{p_0'}^{\circ} \cdot \overline{p_0'} \cdots w_i \cdot \overline{p_i'}^{\circ} \cdot \overline{p_i'}) = \tilde{s}_i$ for $0 \leq i \leq m$. This shows $\delta(s', w) = \tilde{s}_m = s'$.

Statement 1: By induction hypothesis we have $p_0', \ldots, p_m' \in \mathbb{L}_l^{\mathcal{T}}$. By Lemma 3.8, $\lambda(p_m') \in \mathbb{L}_l^{\mathcal{T}}$. Since $w \in A^+$ and $w_0\overline{p_0'}^{\circ}, \ldots, w_m\overline{p_m'}^{\circ} \in A^+$, it follows that $\pi(p, r) \in \mathrm{IT}(\mathbb{L}_l^{\mathcal{T}}) = \mathbb{L}_{l+1}^{\mathcal{T}}$.

Statement 2: Since p appears at state s, there exist states $\tilde{r}_i, \tilde{s}_i \in S$ with $0 \leq i \leq m$ such that $\tilde{r}_0 = \delta(s, w_0)$, $\tilde{s}_m = s$, $\delta(\tilde{s}_j, w_{j+1}) = \tilde{r}_{j+1}$ for $0 \leq j < m$ and the states $\tilde{r}_i, \tilde{s}_i$ are connected via p_i for

$0 \le i \le m$. Using the additional states $\tilde{r}_j =_{\text{def}} s$ and $\tilde{s}_j =_{\text{def}} s$ for $m+1 \le j \le m+r!-1$ we will show that also p' appears at state s. With $m' =_{\text{def}} m + r! - 1$ it suffices to show the following:

2a. $\tilde{r}_0 = \delta(s, w_0 \cdot \overline{p'_0}^{\circ})$

2b. $\tilde{s}_{m'} = s$

2c. $\delta(\tilde{s}_j, w_{j+1} \cdot \overline{p'_{j+1}}^{\circ}) = \tilde{r}_{j+1}$ for $0 \le j < m$

2d. $\delta(\tilde{s}_j, w) = s = \tilde{r}_{j+1}$ for $m \le j < m'$

2e. the states $\tilde{r}_i, \tilde{s}_i$ are connected via p'_i for $0 \le i \le m$

2f. the states $\tilde{r}_i, \tilde{s}_i$ are connected via $\lambda(p'_m)$ for $m+1 \le i \le m'$

We already know that $\delta(s, w_0) = \tilde{r}_0$ and that the states $\tilde{r}_0, \tilde{s}_0$ are connected via p_0. From Lemma 3.6.1 it follows that p_0 appears at state $\tilde{r}_0$. By induction hypothesis p'_0 appears at state $\tilde{r}_0$. From Lemma 3.6.3 we conclude $\delta(\tilde{r}_0, \overline{p'_0}^{\circ}) = \tilde{r}_0$. This shows 2a. Statement 2b follows from the definition of $\tilde{s}_{m'}$. Now choose some j with $0 \le j < m$. By assumption we have $\delta(\tilde{s}_j, w_{j+1}) = \tilde{r}_{j+1}$ and the states $\tilde{r}_{j+1}, \tilde{s}_{j+1}$ are connected via p_{j+1}. Thus p_{j+1} appears at state $\tilde{r}_{j+1}$ (Lemma 3.6.1). From the induction hypothesis it follows that p'_{j+1} appears at state $\tilde{r}_{j+1}$. We obtain $\delta(\tilde{r}_{j+1}, \overline{p'_{j+1}}^{\circ}) = \tilde{r}_{j+1}$ (Lemma 3.6.3), this shows statement 2c. Note that $\tilde{s}_i = s$ for $m \le i \le m'$ and $\tilde{r}_j = s$ for $m+1 \le j \le m'$. Therefore, statement 2d follows, since we have already seen that $\delta(s, w) = s$. By assumption, the states $\tilde{r}_i, \tilde{s}_i$ are connected via p_i for $0 \le i \le m$. From the induction hypothesis it follows that $\tilde{r}_i, \tilde{s}_i$ are also connected via p'_i for $0 \le i \le m$. This shows statement 2e. From Lemma 3.6.5 it follows that p_m appears at state s. From induction hypothesis it follows that p'_m appears at state s. Thus the states s, s are connected via $\lambda(p'_m)$ by Lemma 3.8. This shows statement 2f, and we have finished the induction step for statement 2 of the lemma.

Statement 3: By assumption, the states s_1, s_2 are connected via p. It follows that p appears at s_1 and at s_2, and there exist states $\tilde{s}_i \in S$ with $0 \le i \le m$ such that $\tilde{s}_0 = \delta(s_1, w_0)$, $\tilde{s}_m = s_2$, $\delta(\tilde{s}_j, w_{j+1}) = \tilde{s}_{j+1}$ for $0 \le j < m$ and p_i appears at state $\tilde{s}_i$ for $0 \le i \le m$. From statement 2 it follows that p' appears at s_1 and at s_2. Let $m' =_{\text{def}} m + r! - 1$ and $\tilde{s}_j =_{\text{def}} s_2$ for $m+1 \le j \le m'$. The states $\tilde{s}_j$ with $0 \le j \le m'$ will witness that s_1, s_2 are connected via p'. To this end it suffices to show the following.

3a. $\tilde{s}_0 = \delta(s_1, w_0 \cdot \overline{p'_0}^{\circ})$

3b. $\tilde{s}_{m'} = s_2$

3c. $\delta(\tilde{s}_j, w_{j+1} \cdot \overline{p'_{j+1}}^{\circ}) = \tilde{s}_{j+1}$ for $0 \le j < m$

3d. $\delta(\tilde{s}_j, w) = \tilde{s}_{j+1}$ for $m \le j < m'$

3e. p'_i appears at state $\tilde{s}_i$ for $0 \le i \le m$

3f. $\lambda(p'_m)$ appears at state $\tilde{s}_j$ for $m < j \le m'$

By assumption, $\delta(s_1, w_0) = \tilde{s}_0$ and p_0 appears at $\tilde{s}_0$. From induction hypothesis it follows that also p'_0 appears at $\tilde{s}_0$. With Lemma 3.6.3 we obtain $\delta(\tilde{s}_0, \overline{p'_0}^{\circ}) = \tilde{s}_0$. This shows statement 3a. Note that 3b is by definition. For $0 \le j < m$ we have $\delta(\tilde{s}_j, w_{j+1}) = \tilde{s}_{j+1}$ by assumption. Since p_{j+1} appears at $\tilde{s}_{j+1}$, it follows that also p'_{j+1} appears at $\tilde{s}_{j+1}$ (induction hypothesis) and $\delta(\tilde{s}_{j+1}, \overline{p'_{j+1}}^{\circ}) = \tilde{s}_{j+1}$ (Lemma 3.6.3). This implies statement 3c. At the beginning of the induction step we have already seen that $\delta(s', w) = s'$ for all $s' \in S$ where p appears. It follows that $\delta(s_2, w) = s_2$. This shows statement 3d, since $\tilde{s}_j = s_2$ for $m \le j \le m'$. We know that p_i appears at $\tilde{s}_i$ for $0 \le i \le m$. Thus p'_i appears at $\tilde{s}_i$ (induction hypothesis) and statement 3e follows. Particularly, p'_m appears at $\tilde{s}_m = s_2 = \tilde{s}_j$ for $m < j \leq m'$. Together with the Lemmas 3.8 and 3.6.1 this shows statement 3f. Thus we have proven statement 3.

Statement 4: Let F' be a dfa with $|F'| \leq r$. We will show that $w^{r!-1}$ leads to states in F' where p' appears. Assume for the moment that we have already shown this, then we argue as follows. First of all we observe that the following holds.

$$
\begin{aligned}
\overline{p'} &= w_0 \cdot \overline{p'_0}^{\circ} \cdots w_m \cdot \overline{p'_m}^{\circ} \cdot w^{r!-1} \\
\overline{p'}^{\circ} &= w_0 \cdot \overline{p'_0}^{\circ} \cdot \overline{p'_0} \cdots w_m \cdot \overline{p'_m}^{\circ} \cdot \overline{p'_m} \cdot \left(w \cdot \overline{\lambda(p'_m)}\right)^{r!-1}
\end{aligned}
$$

Hence $\overline{p'}$ leads to states in F' where p' appears. Note that $\overline{p'_m}$ is a suffix of w. From induction hypothesis it follows that w leads to states in F' where p'_m appears. From the Lemmas 3.8 and 3.6.2 we obtain that w leads to a $\overline{\lambda(p'_m)}$-loop in F'. Hence $w \cdot \overline{\lambda(p'_m)} \sim_{F'} w$ and $\overline{p'}^{\circ} \sim_{F'} w_0 \cdot \overline{p'_0}^{\circ} \cdot \overline{p'_0} \cdots w_m \cdot \overline{p'_m}^{\circ} \cdot \overline{p'_m} \cdot w^{r!-1}$. It follows that $\overline{p'}^{\circ}$ leads to states in F' where p' appears. This implies statement 4 and it remains to show that $w^{r!-1}$ leads to states in F' where p' appears.

By Proposition 2.4, $w^{r!-1} = w^{r!-r-1} \cdot w^r$ leads to a $w^{r!}$-loop in F'. Thus it suffices to show that p' appears at s' for every state $s' \in S$ with $s' = \delta(s', w^{r!})$. For this purpose let $m' =_{\text{def}} m + r! - 1$ and define the following states.

$$
\begin{aligned}
\tilde{r}_0 &=_{\text{def}} \delta(s', w_0 \cdot \overline{p'_0}^{\circ}) \\
\tilde{s}_0 &=_{\text{def}} \delta(\tilde{r}_0, \overline{p'_0}) \\
\tilde{r}_i &=_{\text{def}} \delta(\tilde{s}_{i-1}, w_i \cdot \overline{p'_i}^{\circ}) \quad \text{for } 1 \leq i \leq m \\
\tilde{s}_i &=_{\text{def}} \delta(\tilde{r}_i, \overline{p'_i}) \quad \text{for } 1 \leq i \leq m \\
\tilde{r}_{m+j} &=_{\text{def}} \delta(\tilde{s}_m, w^j) \quad \text{for } 1 \leq j \leq r! - 1 \\
\tilde{s}_{m+j} &=_{\text{def}} \tilde{r}_{m+j} \quad \text{for } 1 \leq j \leq r! - 1
\end{aligned}
$$

We want to show that these are the states witnessing that p' appears at state s'. It is easy to see that $\delta(s', w_0 \cdot \overline{p'_0}^{\circ} \cdot \overline{p'_0} \cdots w_i \cdot \overline{p'_i}^{\circ}) = \tilde{r}_i$ and $\delta(s', w_0 \cdot \overline{p'_0}^{\circ} \cdot \overline{p'_0} \cdots w_i \cdot \overline{p'_i}^{\circ} \cdot \overline{p'_i}) = \tilde{s}_i$ for $0 \leq i \leq m$. Since $w = w_0 \cdot \overline{p'_0}^{\circ} \cdot \overline{p'_0} \cdots w_m \cdot \overline{p'_m}^{\circ} \cdot \overline{p'_m}$, we have $\tilde{s}_m = \delta(s', w)$ and $\tilde{r}_{m+j} = \tilde{s}_{m+j} = \delta(\tilde{s}_m, w^j) = \delta(s', w^{j+1})$ for $1 \leq j \leq r! - 1$. It follows that $\tilde{s}_{m'} = \delta(s', w^{r!}) = s'$. To see that p' appears at s', it remains to show the following.

- the states $\tilde{r}_i, \tilde{s}_i$ are connected via p'_i for $0 \leq i \leq m$
- the states $\tilde{r}_j, \tilde{s}_j$ are connected via $\lambda(p'_m)$ for $m < j \leq m'$

The first item follows from the induction hypothesis, since $\overline{p'_i}^{\circ}, \overline{p'_i}^{\circ}\overline{p'_i}$ lead to states which are connected via p'_i (by definition $\tilde{r}_i, \tilde{s}_i$ are such states). Note that $\overline{p'_m}$ is a suffix of w. From induction hypothesis it follows that w leads to states in F' where p'_m appears. Therefore, p'_m appears at $\tilde{r}_j = \tilde{s}_j$ for $m < j \leq m'$. From Lemma 3.8 it follows that $\tilde{r}_j, \tilde{s}_j$ are connected via $\lambda(p'_m)$. This shows the second item. Hence we have proven statement 4 of the lemma.

Statement 5: Let F' be some dfa with $|F'| \leq r$ and let s' be an arbitrary state of F'. For $s'_1 =_{\text{def}} \delta(s', \overline{p'}^{\circ})$ and $s'_2 =_{\text{def}} \delta(s', \overline{p'}^{\circ}\overline{p'})$ we want to show that s'_1, s'_2 are connected via p'. For this let $m' =_{\text{def}} m + r! - 1$ and define the following witnessing states.

$$
\begin{aligned}
\tilde{s}_0 &=_{\text{def}} \delta(s'_1, w_0 \cdot \overline{p'_0}^{\circ}) && (8) \\
\tilde{s}_{i+1} &=_{\text{def}} \delta(\tilde{s}_i, w_{i+1} \cdot \overline{p'_{i+1}}^{\circ}) \quad \text{for } 0 \leq i < m && (9) \\
\tilde{s}_{j+1} &=_{\text{def}} \delta(\tilde{s}_j, w) \quad \text{for } m \leq j < m' && (10)
\end{aligned}
$$

By statement 4, p' appears at s'_1 and at s'_2. Furthermore, it is easy to see that $\tilde{s}_{m'} = \delta(s'_1, \overline{p'}) = s'_2$. It remains to show that (i) p'_i appears at state $\tilde{s}_i$ for $0 \leq i \leq m$ and (ii) $\lambda(p'_m)$ appears at state $\tilde{s}_j$ for

$m+1 \le j \le m'$. By induction hypothesis, for $0 \le i \le m$ it holds that $\overline{p'_i}^{\circ}$ leads to states in F' where p'_i appears. From the equations (8) and (9) it follows that p'_i appears at state $\tilde{s}_i$ for $0 \le i \le m$. From induction hypothesis it follows that $\overline{p'_m}$ leads to states in F' where p'_m appears. Hence also w leads to states in F' where p'_m appears (note that $\overline{p'_m}$ is a suffix of w). Thus from equation (10) we obtain that p'_m appears at state $\tilde{s}_j$ for $m+1 \le j \le m'$. From Lemma 3.8 it follows that $\tilde{s}_j, \tilde{s}_j$ are connected via $\lambda(p'_m)$ for $m+1 \le j \le m'$. Particularly, $\lambda(p'_m)$ appears at state $\tilde{s}_j$ for $m+1 \le j \le m'$ (Lemma 3.6.1). This shows statement 5.

Statement 6: This can be seen analogously to the proof of statement 5. There we only have to replace the terms "$\overline{p'}^{\circ}$" by "$\overline{p'}$". ❑

The following lemma isolates the main argument of the proof of Theorem 3.15 which establishes a relation between the polynomial closure operation and the iteration rule for the patterns classes. The lemma says that under certain assumptions we can replace bridge–words by their respective loop–words without leaving the language of some dfa.

Lemma 3.11. *Let $\mathcal{T}$ be an initial pattern, $n \ge 0$, $r \ge 3$, $p \in \mathbb{L}^{\mathcal{T}}_{n+1}$ and $p' =_{\mathrm{def}} \pi(p,r)$. Furthermore, let $F = (A, S, \delta, s_0, S')$ with $|F| \le r$ be a dfa which does not have pattern $\mathbb{L}^{\mathcal{T}}_{n}$. Then for all $u, z \in A^*$ we have*

$$u\overline{p'}z \in L(F) \Longrightarrow u\overline{p'}^{\circ}z \in L(F).$$

Proof. We choose suitable $m \ge 0$, $w_0, \ldots, w_m \in A^+$ and $p_0, \ldots, p_m \in \mathbb{L}^{\mathcal{T}}_n$ such that $p = (w_0, p_0, \ldots, w_m, p_m)$. For $0 \le i \le m$ let $p'_i =_{\mathrm{def}} \pi(p_i, r)$ and $w =_{\mathrm{def}} w_0 \cdot \overline{p'_0}^{\circ} \cdot \overline{p'_0} \cdots w_m \cdot \overline{p'_m}^{\circ} \cdot \overline{p'_m}$. From Definition 3.9 it follows that $p' = (w_0 \cdot \overline{p'_0}^{\circ}, p'_0, \ldots, w_m \cdot \overline{p'_m}^{\circ}, p'_m, w, \lambda(p'_m), \ldots, w, \lambda(p'_m))$ where the term "$w, \lambda(p'_m)$" is repeated $(r!-1)$ times.
Now let $u, z \in A^*$ such that $u\overline{p'}z \in L(F)$. Thus we have

$$\underbrace{uw_0}_{u' =_{\mathrm{def}}} \overline{p'_0}^{\circ} \underbrace{w_1\overline{p'_1}^{\circ} w_2\overline{p'_2}^{\circ} \cdots w_m\overline{p'_m}^{\circ} w^{r!-1} z}_{z' =_{\mathrm{def}}} \in L(F).$$

We want to show $u'\overline{p'_0}^{\circ}\overline{p'_0}z' \in L(F)$. From Lemma 3.10.5 it follows that the states $s_1 =_{\mathrm{def}} \delta(s_0, u'\overline{p'_0}^{\circ})$ and $s_2 =_{\mathrm{def}} \delta(s_0, u'\overline{p'_0}^{\circ}\overline{p'_0})$ are connected via p'_0. Note that $p'_0 \in \mathbb{L}^{\mathcal{T}}_n$ by Lemma 3.10.1. If $u'\overline{p'_0}^{\circ}\overline{p'_0}z' \notin L(F)$ then we have $\delta(s_0, u'\overline{p'_0}^{\circ}) = s_1$, $\delta(s_1, z') \in S'$, $\delta(s_2, z') \notin S'$ and the states s_1, s_2 are connected via $p'_0 \in \mathbb{L}^{\mathcal{T}}_n$. It follows that F has pattern $\mathbb{L}^{\mathcal{T}}_n$. This is a contradiction to the assumption. Thus starting from

$$uw_0\overline{p'_0}^{\circ} \quad w_1\overline{p'_1}^{\circ} w_2\overline{p'_2}^{\circ} \cdots w_m\overline{p'_m}^{\circ} w^{r!-1} z \in L(F)$$

we have shown

$$uw_0\overline{p'_0}^{\circ}\overline{p'_0} w_1\overline{p'_1}^{\circ} w_2\overline{p'_2}^{\circ} \cdots w_m\overline{p'_m}^{\circ} w^{r!-1} z \in L(F).$$

Analogously we obtain:

$$\begin{array}{llllllll}
uw_0\overline{p'_0}^{\circ}\overline{p'_0}w_1\overline{p'_1}^{\circ} & w_2\overline{p'_2}^{\circ} & w_3\overline{p'_3}^{\circ} & w_4\overline{p'_4}^{\circ} & \cdots & w_m\overline{p'_m}^{\circ} & w^{r!-1}z \in L(F) \\
uw_0\overline{p'_0}^{\circ}\overline{p'_0}w_1\overline{p'_1}^{\circ}\overline{p'_1}w_2\overline{p'_2}^{\circ} & & w_3\overline{p'_3}^{\circ} & w_4\overline{p'_4}^{\circ} & \cdots & w_m\overline{p'_m}^{\circ} & w^{r!-1}z \in L(F) \\
uw_0\overline{p'_0}^{\circ}\overline{p'_0}w_1\overline{p'_1}^{\circ}\overline{p'_1}w_2\overline{p'_2}^{\circ}\overline{p'_2}w_3\overline{p'_3}^{\circ} & & & w_4\overline{p'_4}^{\circ} & \cdots & w_m\overline{p'_m}^{\circ} & w^{r!-1}z \in L(F) \\
uw_0\overline{p'_0}^{\circ}\overline{p'_0}w_1\overline{p'_1}^{\circ}\overline{p'_1}w_2\overline{p'_2}^{\circ}\overline{p'_2}w_3\overline{p'_3}^{\circ}\overline{p'_3}w_4\overline{p'_4}^{\circ} & & & & \cdots & w_m\overline{p'_m}^{\circ} & w^{r!-1}z \in L(F) \\
 & & & \vdots & & & \\
uw_0\overline{p'_0}^{\circ}\overline{p'_0}w_1\overline{p'_1}^{\circ}\overline{p'_1}w_2\overline{p'_2}^{\circ}\overline{p'_2}w_3\overline{p'_3}^{\circ}\overline{p'_3}w_4\overline{p'_4}^{\circ} & & & & \cdots & w_m\overline{p'_m}^{\circ}\overline{p'_m} & w^{r!-1}z \in L(F)
\end{array} \tag{11}$$

By definition, $\overline{p'_m}$ is a suffix of w. From Lemma 3.10.4 it follows that w leads to states in F where p'_m appears. From Lemma 3.8 we obtain that for all $s' \in S$ with $s =_{\mathrm{def}} \delta(s', w)$ it holds that s, s

are connected via $\lambda(p'_m)$. Now from Lemma 3.6.2 it follows that w leads to states in F which have a $\overline{\lambda(p'_m)}$-loop. Thus from equation (11) we obtain

$$uw_0\overline{p'_0}^{\circ}\,\overline{p'_0}w_1\overline{p'_1}^{\circ}\,\overline{p'_1}w_2\overline{p'_2}^{\circ}\,\overline{p'_2}w_3\overline{p'_3}^{\circ}\,\overline{p'_3}w_4\overline{p'_4}^{\circ}\,\overline{p'_4}\cdots w_m\overline{p'_m}^{\circ}\,\overline{p'_m}\;w\overline{\lambda(p'_m)}\;w^{r!-2}z \in L(F).$$

We can repeat this argument for the remaining $(r!-2)$ occurrences of w to see that

$$u\,\underbrace{w_0\overline{p'_0}^{\circ}\,\overline{p'_0}w_1\overline{p'_1}^{\circ}\,\overline{p'_1}w_2\overline{p'_2}^{\circ}\,\overline{p'_2}w_3\overline{p'_3}^{\circ}\,\overline{p'_3}w_4\overline{p'_4}^{\circ}\,\overline{p'_4}\cdots w_m\overline{p'_m}^{\circ}\,\overline{p'_m}w\overline{\lambda(p'_m)}\ldots w\overline{\lambda(p'_m)}}_{=\overline{p'}^{\circ}}\,z \in L(F)$$

where the term "$w, \overline{\lambda(p'_m)}$" is repeated $(r!-1)$ times. This shows $u\overline{p'}^{\circ}z \in L(F)$. ❑

In forthcoming proofs we make use of the fact that the structure of arbitrary patterns can be found at each sink of some dfa. More precisely we can state the following proposition.

Proposition 3.12. *Let $F = (A, S, \delta, s_0, S')$ be a dfa and suppose there is some $s \in S$ such that $\delta(s, a) = s$ for all $a \in A$. Let $\mathcal{T}$ be an initial pattern, let $n \geq 0$ and $p \in \mathbb{L}_n^{\mathcal{T}}$. Then p appears at s, and s, s are connected via p.*

Proof. The proof is by induction on n. The induction base for $n = 0$ is trivial, just consider Definition 3.4. Now suppose the proposition holds for some $n \geq 0$ and we want to show it for $n+1$. So let $p = (w_0, p_0, \ldots, w_m, p_m) \in \mathbb{L}_{n+1}^{\mathcal{T}}$ for some $m \geq 0$ with $w_i \in A^+$ and $p_i \in \mathbb{L}_n^{\mathcal{T}}$. By hypothesis, we know for $0 \leq i \leq m$ that p_i appears at s, and s and s are connected via p_i. If we set the required states in Definition 3.4 all to s it is easy to verify with help of the hypothesis that p appears at s, and s and s are connected via p. Just observe that $\delta(s, w_i) = s$ for $0 \leq i \leq m$. ❑

3.3 Pattern Iterator versus Polynomial Closure

The main result in this section is stated in Theorem 3.15: A complementation followed by a polynomial closure operation on the language side is captured by our iteration rule on the forbidden pattern side. To prove this we make use of the auxiliary results given in subsection 3.2 where Lemma 3.11 contains the main argument.

First of all we use our hierarchies of pattern classes to define hierarchies of language classes via a forbidden pattern approach. So each initial pattern induces a hierarchy of language classes. We treat the inclusion structure and the decidability question of these hierarchies in the subsections 3.4 and 3.6.

Definition 3.13. *Let $\mathcal{T}$ be an initial pattern. For $n \geq 0$ we define the class of languages corresponding to $\mathbb{L}_n^{\mathcal{T}}$ as $\mathcal{C}_n^{\mathcal{T}} =_{\text{def}} \{L \subseteq A^+ :\ L$ is accepted by some dfa F which does not have pattern $\mathbb{L}_n^{\mathcal{T}}\}$.*

The following proposition shows that the classes from Definition 3.13 are well defined.

Proposition 3.14. *For $n \geq 0$, an initial pattern $\mathcal{T}$ and dfa's F_1, F_2 with $L(F_1) = L(F_2)$ we have:*

$$F_1 \textit{ has pattern } \mathbb{L}_n^{\mathcal{T}} \iff F_2 \textit{ has pattern } \mathbb{L}_n^{\mathcal{T}}$$

Proof. Suppose F_1 has pattern $\mathbb{L}_n^{\mathcal{T}}$. Then there are states s_1, s_2 in F_1 which are connected via some $p \in \mathbb{L}_n^{\mathcal{T}}$ such that $\delta(s_0, u) = s_1$, $\delta(s_1, z) \in S'$, $\delta(s_2, z) \notin S'$ for suitable $u, z \in A^*$, and for the starting state s_0 of F_1 and the set of accepting states S' of F_1. Define $r =_{\text{def}} |F_2|$ and $p' =_{\text{def}} \pi(p, r)$. We obtain from Lemma 3.10.3 that s_1, s_2 in F_1 are also connected via p'. So $u\overline{p'}^{\circ}z \in L(F_1) = L(F_2)$ (Lemma 3.6.3) and $u\overline{p'}^{\circ}\overline{p'}z \notin L(F_1) = L(F_2)$ (Lemma 3.6.2). Now define s'_1 and s'_2 to be the states in F_2 that can be reached from its starting state by $u\overline{p'}^{\circ}$ and $u\overline{p'}^{\circ}\overline{p'}$, respectively. By Lemma 3.10.5 we get that s'_1 and s'_2 are connected via p' in F_2. Since we reach from s'_1 (s'_2) with z an accepting state (rejecting state, respectively) of F_2, this shows that also F_2 has pattern $\mathbb{L}_n^{\mathcal{T}}$.

If we change the roles of F_1 and F_2 the same arguments also show the other implication. ❑

Theorem 3.15. *Let $\mathcal{T}$ be an initial pattern and $n \geq 0$, then* $\mathrm{Pol}(\mathrm{co}\mathcal{C}_n^{\mathcal{T}}) \subseteq \mathcal{C}_{n+1}^{\mathcal{T}}$.

Proof. Let $\mathcal{T}$ be an initial pattern and $n \geq 0$. We assume that there exists an $L \in \mathrm{Pol}(\mathrm{co}\mathcal{C}_n^{\mathcal{T}})$ which is no element of $\mathcal{C}_{n+1}^{\mathcal{T}}$, this will lead to a contradiction. From $L \in \mathrm{Pol}(\mathrm{co}\mathcal{C}_n^{\mathcal{T}})$ it follows that

$$L = \bigcup_{i=1}^{k} L_{i,0} L_{i,1} \cdots L_{i,k_i}$$

for some $k \geq 0$, $k_i \geq 0$ and $L_{i,j} \in \mathrm{co}\mathcal{C}_n^{\mathcal{T}}$. Let $F = (A, S, \delta, s_0, S')$ be a dfa with $L(F) = L$. For $1 \leq i \leq k$ and $0 \leq j \leq k_i$ let $F_{i,j}$ be a dfa with $L(F_{i,j}) = L_{i,j}$ and let $F'_{i,j}$ be a dfa with $L(F'_{i,j}) = A^+ \setminus L_{i,j}$. Furthermore, we define

$$r =_{\mathrm{def}} \max\left(\left\{|F_{i,j}|, |F'_{i,j}| : 1 \leq i \leq k \wedge 0 \leq j \leq k_i\right\} \cup \{|F|\} \cup \{k_i + 1 : 1 \leq i \leq k\}\right).$$

The dfa F has pattern $\mathbb{L}_{n+1}^{\mathcal{T}}$, since $L \subseteq A^+$ and $L \notin \mathcal{C}_{n+1}^{\mathcal{L}}$ (cf. Definition 3.13 and Proposition 3.14). So there exist $s_1, s_2 \in S$, $u, z \in A^*$, $p \in \mathbb{L}_{n+1}^{\mathcal{T}}$ such that $\delta(s_0, u) = s_1$, $\delta(s_1, z) \in S'$, $\delta(s_2, z) \notin S'$ and the states s_1, s_2 are connected via p. It follows that $L \neq \emptyset$ and $k \geq 1$. By Lemma 3.10.2, the states s_1, s_2 are also connected via pattern $p' =_{\mathrm{def}} \pi(p, r)$. From Lemma 3.6.3 it follows that $u\big(\overline{p'}^{\circ}\big)^i z \in L$ for all $i \geq 0$. Thus there exists an i' with $1 \leq i' \leq k$ such that

$$u\left(\overline{p'}^{\circ}\right)^r z \in L_{i',0} L_{i',1} \cdots L_{i',k_{i'}}.$$

Since $r \geq k_{i'} + 1$, it follows that there exist $0 \leq j' \leq k_{i'}$ and words $u', u'', z', z'' \in A^*$ such that:

1. $u\left(\overline{p'}^{\circ}\right)^r z = u''u'\overline{p'}^{\circ} z' z''$
2. $u''u' = u\left(\overline{p'}^{\circ}\right)^i$ and $z'z'' = \left(\overline{p'}^{\circ}\right)^j z$ for some $i, j \geq 0$
3. $u'' \in L_{i',0} L_{i',1} \cdots L_{i',j'-1}$, $\quad u'\overline{p'}^{\circ} z' \in L_{i',j'}$, $\quad z'' \in L_{i',j'+1} L_{i',j'+2} \cdots L_{i',k_{i'}}$

Since $|F_{i',j'}| \leq r$, the word $\overline{p'}^{\circ}$ leads to states in $F_{i',j'}$ where p' appears (Lemma 3.10.4). Particularly, such a state has a $\overline{p'}^{\circ}$-loop (Lemma 3.6.3). From $u'\overline{p'}^{\circ} z' \in L_{i',j'}$ it follows that for all $i \geq 1$ we have

$$u'\left(\overline{p'}^{\circ}\right)^i z' \in L_{i',j'}. \tag{12}$$

Since s_1, s_2 are connected via p' in F, we have $\delta(s_0, u''u') = s_1$, $\delta(s_2, z'z'') = \delta(s_2, z)$ and $\delta(s_1, \overline{p'}) = s_2$ (Lemma 3.6.2). Assume that $u'\overline{p'}^{\circ}\overline{p'}z' \in L_{i',j'}$, then we would obtain $u''u'\overline{p'}^{\circ}\overline{p'}z'z'' \in L$. It would follow that

$$\delta(s_2, z) = \delta(s_2, z'z'') = \delta(s_1, \overline{p'}z'z'') = \delta(s_1, \overline{p'}^{\circ}\overline{p'}z'z'') = \delta(s_0, u''u'\overline{p'}^{\circ}\overline{p'}z'z'') \in S'.$$

This is a contradiction since $\delta(s_2, z) \notin S'$. It follows that

$$u'\overline{p'}^{\circ}\overline{p'}z' \notin L_{i',j'}. \tag{13}$$

From $u'\overline{p'}^{\circ} z' \in L_{i',j'} \subseteq A^+$, it follows that $0 < |u'\overline{p'}^{\circ} z'| \leq |u'\overline{p'}^{\circ}\overline{p'}z'|$. From equation (13) we obtain

$$u'\overline{p'}^{\circ}\overline{p'}z' \in A^+ \setminus L_{i',j'} = L(F'_{i',j'}). \tag{14}$$

Observe that $L(F'_{i',j'}) \in \mathcal{C}_n^{\mathcal{T}}$. Therefore, dfa $F'_{i',j'}$ does not have pattern $\mathbb{L}_n^{\mathcal{T}}$. Since $|F'_{i',j'}| \leq r$ we can apply Lemma 3.11. From equation (14) we obtain $u'\overline{p'}^{\circ}\overline{p'}^{\circ} z' \in L(F'_{i',j'})$. It follows that $u'\overline{p'}^{\circ}\overline{p'}^{\circ} z' \notin A^+ \setminus L(F'_{i',j'}) = L_{i',j'}$. This is a contradiction to equation (12). It follows that $\mathrm{Pol}(\mathrm{co}\mathcal{C}_n^{\mathcal{T}}) \subseteq \mathcal{C}_{n+1}^{\mathcal{T}}$. ❑

3.4 Inclusion Structure of the classes $\mathcal{C}_n^{\mathcal{T}}$

In this subsection we show that if some initial pattern $\mathcal{T}$ satisfies a certain (weak) property then the inclusion $\mathcal{C}_n^{\mathcal{T}} \cup \mathrm{co}\mathcal{C}_n^{\mathcal{T}} \subseteq \mathcal{C}_{n+1}^{\mathcal{T}} \cap \mathrm{co}\mathcal{C}_{n+1}^{\mathcal{T}}$ holds for all $n \geq 0$.

Definition 3.16. *For initial patterns $\mathcal{T}_1, \mathcal{T}_2$ and $n_1, n_2 \geq 0$ we define $\mathbb{L}_{n_1}^{\mathcal{T}_1} \preceq \mathbb{L}_{n_2}^{\mathcal{T}_2}$ if and only if for every $p_2 \in \mathbb{L}_{n_2}^{\mathcal{T}_2}$ there exists an $p_1 \in \mathbb{L}_{n_1}^{\mathcal{T}_1}$ such that for every dfa $F = (A, S, \delta, s_0, S')$ and all states $s, s_1, s_2 \in S$ the following holds:*

1. *If p_2 appears at s, then p_1 appears at s.*
2. *If s_1, s_2 are connected via p_2, then s_1, s_2 are connected via p_1.*

Lemma 3.17. *For initial patterns $\mathcal{T}_1, \mathcal{T}_2$ and $n_1, n_2 \geq 0$ the following holds.*

$$\mathbb{L}_{n_1}^{\mathcal{T}_1} \preceq \mathbb{L}_{n_2}^{\mathcal{T}_2} \Longrightarrow \mathbb{L}_{n_1+1}^{\mathcal{T}_1} \preceq \mathbb{L}_{n_2+1}^{\mathcal{T}_2}$$

Proof. Let $\mathcal{T}_1, \mathcal{T}_2$ be initial patterns and $n_1, n_2 \geq 0$ such that $\mathbb{L}_{n_1}^{\mathcal{T}_1} \preceq \mathbb{L}_{n_2}^{\mathcal{T}_2}$. Hence for a given $p_2 = (w_{2,0}, p_{2,0}, \ldots, w_{2,m}, p_{2,m}) \in \mathbb{L}_{n_2+1}^{\mathcal{T}_2}$ there exist $p_{1,0}, \ldots, p_{1,m} \in \mathbb{L}_{n_1}^{\mathcal{T}_1}$ such that for every dfa $F = (A, S, \delta, s_0, S')$ and all states $s, s_1, s_2 \in S$ the following holds:

$$\text{If } p_{2,i} \text{ appears at } s, \text{ then } p_{1,i} \text{ appears at } s \text{ for } 0 \leq i \leq m. \tag{15}$$

$$\text{If } s_1, s_2 \text{ are connected via } p_{2,i}, \text{ then } s_1, s_2 \text{ are connected via } p_{1,i} \text{ for } 0 \leq i \leq m. \tag{16}$$

Define $p_1 =_{\mathrm{def}} (w_{2,0}, p_{1,0}, \ldots, w_{2,m}, p_{1,m})$ and observe that $p_1 \in \mathbb{L}_{n_1+1}^{\mathcal{T}_1}$. Now let $F = (A, S, \delta, s_0, S')$ be a dfa and $s, s_1, s_2 \in S$. We want to show the following:

(i) If p_2 appears at s, then p_1 appears at s.

(ii) If s_1, s_2 are connected via p_2, then s_1, s_2 are connected via p_1.

Suppose that p_2 appears at s, then there are states $\tilde{r}_i, \tilde{s}_i \in S$ with $0 \leq i \leq m$ such that $\tilde{r}_0 = \delta(s, w_{2,0})$, $\tilde{s}_m = s$, $\delta(\tilde{s}_j, w_{2,j+1}) = \tilde{r}_{j+1}$ for $0 \leq j < m$ and the states $\tilde{r}_i, \tilde{s}_i$ are connected via $p_{2,i}$ for $0 \leq i \leq m$. From (16) it follows that the states $\tilde{r}_i, \tilde{s}_i$ are also connected via $p_{1,i}$ for $0 \leq i \leq m$. Therefore, p_1 appears at s, and we have shown statement (i).

Suppose now that s_1, s_2 are connected via p_2. By definition, p_2 appears at the states s_1, s_2 and there exist states $\tilde{s}_i \in S$ with $0 \leq i \leq m$ such that $\tilde{s}_0 = \delta(s_1, w_{2,0})$, $\tilde{s}_m = s_2$, $\delta(\tilde{s}_j, w_{2,j+1}) = \tilde{s}_{j+1}$ for $0 \leq j < m$ and $p_{2,i}$ appears at state $\tilde{s}_i$ for $0 \leq i \leq m$. From statement (i) we obtain that p_1 appears at state s_1 and at state s_2. Furthermore, from (15) it follows that also $p_{1,i}$ appears at state $\tilde{s}_i$ for $0 \leq i \leq m$. Hence s_1, s_2 are connected via p_1, and we have proven statement (ii).

It follows that $\mathbb{L}_{n_1+1}^{\mathcal{T}_1} \preceq \mathbb{L}_{n_2+1}^{\mathcal{T}_2}$. This proves the lemma. ❑

Lemma 3.18. *For initial patterns $\mathcal{T}_1, \mathcal{T}_2$ and $n_1, n_2 \geq 0$ the following holds.*

$$\mathbb{L}_{n_1}^{\mathcal{T}_1} \preceq \mathbb{L}_{n_2}^{\mathcal{T}_2} \Longrightarrow \mathcal{C}_{n_1}^{\mathcal{T}_1} \subseteq \mathcal{C}_{n_2}^{\mathcal{T}_2}$$

Proof. Let $\mathcal{T}_1, \mathcal{T}_2$ be initial patterns and $n_1, n_2 \geq 0$ such that $\mathbb{L}_{n_1}^{\mathcal{T}_1} \preceq \mathbb{L}_{n_2}^{\mathcal{T}_2}$. For an arbitrary language $L \subseteq A^+$ with $L \notin \mathcal{C}_{n_2}^{\mathcal{T}_2}$ we want to show that $L \notin \mathcal{C}_{n_1}^{\mathcal{T}_1}$. Since $L \notin \mathcal{C}_{n_2}^{\mathcal{T}_2}$, the dfa $F = (A, S, \delta, s_0, S')$ with $L(F) = L$ has pattern $\mathbb{L}_{n_2}^{\mathcal{T}_2}$. Thus there exist $s_1, s_2 \in S$, $u, z \in A^*$, $p_2 \in \mathbb{L}_{n_2}^{\mathcal{T}_2}$ such that $\delta(s_0, u) = s_1$, $\delta(s_1, z) \in S'$, $\delta(s_2, z) \notin S'$ and the states s_1, s_2 are connected via p_2. Since $\mathbb{L}_{n_1}^{\mathcal{T}_1} \preceq \mathbb{L}_{n_2}^{\mathcal{T}_2}$, there exists a $p_1 \in \mathbb{L}_{n_1}^{\mathcal{T}_1}$ such that the states s_1, s_2 are connected via p_1. It follows that F has also pattern $\mathbb{L}_{n_1}^{\mathcal{T}_1}$. This shows $L \notin \mathcal{C}_{n_1}^{\mathcal{T}_1}$, and hence $\mathcal{C}_{n_1}^{\mathcal{T}_1} \subseteq \mathcal{C}_{n_2}^{\mathcal{T}_2}$. ❑

Lemma 3.19. *For an initial pattern $\mathcal{T}$ and $n \geq 0$ the following holds.*

1. $\mathrm{co}\mathcal{C}_n^{\mathcal{T}} \subseteq \mathcal{C}_{n+1}^{\mathcal{T}}$
2. $\mathcal{C}_n^{\mathcal{T}} \subseteq \mathrm{co}\mathcal{C}_{n+1}^{\mathcal{T}}$

Proof. From Theorem 3.15 it follows that $\mathrm{co}\mathcal{C}_n^{\mathcal{T}} \subseteq \mathrm{Pol}(\mathrm{co}\mathcal{C}_n^{\mathcal{T}}) \subseteq \mathcal{C}_{n+1}^{\mathcal{T}}$. This also implies $\mathcal{C}_n^{\mathcal{T}} = \mathrm{co}\,(\mathrm{co}\mathcal{C}_n^{\mathcal{T}}) \subseteq \mathrm{co}\mathcal{C}_{n+1}^{\mathcal{T}}$. ❑

Theorem 3.20. *Let $\mathcal{T}$ be an initial pattern with $\mathbb{L}_0^{\mathcal{T}} \preceq \mathbb{L}_1^{\mathcal{T}}$, then for $n \geq 0$ the following holds.*

$$\mathcal{C}_n^{\mathcal{T}} \cup \mathrm{co}\mathcal{C}_n^{\mathcal{T}} \subseteq \mathcal{C}_{n+1}^{\mathcal{T}} \cap \mathrm{co}\mathcal{C}_{n+1}^{\mathcal{T}}$$

Proof. From Lemma 3.17 we obtain $\mathbb{L}_n^{\mathcal{T}} \preceq \mathbb{L}_{n+1}^{\mathcal{T}}$ for $n \geq 0$. Lemma 3.18 implies $\mathcal{C}_n^{\mathcal{T}} \subseteq \mathcal{C}_{n+1}^{\mathcal{T}}$ for $n \geq 0$. From this we conclude $\mathrm{co}\mathcal{C}_n^{\mathcal{T}} \subseteq \mathrm{co}\mathcal{C}_{n+1}^{\mathcal{T}}$ for $n \geq 0$. Together with Lemma 3.19 this proves the theorem. ❑

3.5 Pattern Iteration remains Starfree

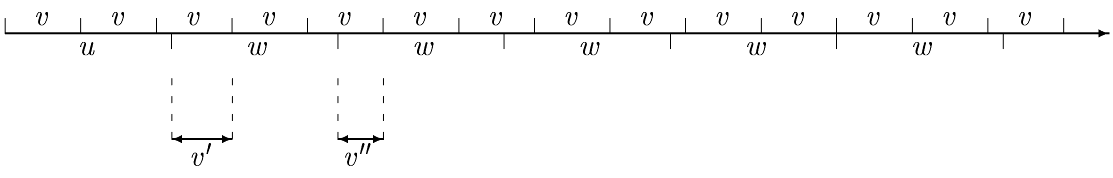
In this subsection we show that the pattern iterator IT can be considered as a starfree iterator. Let $\mathcal{T}$ be an arbitrary initial pattern and recall that SF denotes the class of starfree languages. In Theorem 3.26 we show that for $n \geq 1$ it holds that $\mathcal{C}_n^{\mathcal{T}} \subseteq SF$ if and only if $\bigcup_{i \geq 0} \mathcal{C}_i^{\mathcal{T}} \subseteq SF$. For the proof of this theorem we need some auxiliary results on periodic infinite words (cf. Lemma 3.21) and a modification of Proposition 2.2 (cf. Lemma 3.24) by taking forbidden patterns into account. We also make a remark on the restriction $n \geq 1$ in Theorem 3.26.

In this subsection we will consider infinite words, where we take over certain notions from finite words. If $w \in A^+$ with $w = a_1 \cdots a_m$ for alphabet letters a_i, then w^∞ denotes the infinite word $a_1 \cdots a_m a_1 \cdots a_m \cdots$. Let $0 \leq k \leq l$ and $v \in A^*$ with $v = b_1 \cdots b_l$ for alphabet letters b_i, then $b_{l-k+1} b_{l-k+2} \cdots b_l$ is the k–suffix of v. For $m \geq 0$ and $n \geq 1$ we use $(m \mod n)$ as an abbreviation for $m - n\lfloor m/n \rfloor$.

Lemma 3.21. *Let $v \in A^+$ such that there do not exist $u' \in A^*$, $v' \in A^+$ with $|v'| < |v|$ and $v^\infty = u'v'^\infty$. If $v^\infty = uw^\infty$ for some $u \in A^*$, $w \in A^+$ then $|w|$ is a multiple of $|v|$.*

Proof. Let $v = a_1 \cdots a_m$ with $a_i \in A$ as in the lemma. We assume that $v^\infty = uw^\infty$ for some $u \in A^*$, $w \in A^+$ where $n =_{\mathrm{def}} |w|$ is not a multiple of m. This will lead to a contradiction.

Let $m' =_{\mathrm{def}} m - (|u| \mod m)$ and $m'' =_{\mathrm{def}} m - (|uw| \mod m)$, and note that $1 \leq m', m'' \leq m$. Moreover, let v' be the m'–suffix of v, and let v'' be the m''–suffix of v. Then we have the following decomposition of $v^\infty = uw^\infty$.



Observe that $|v'| \neq |v''|$. Otherwise we would obtain $(|u| \mod m) = (|uw| \mod m)$ which implies that $|w|$ is a multiple of m. This is a contradiction to our assumption.

Note that $uv'v^\infty = uwv''v^\infty = v^\infty$. It follows that $w^\infty = v'v^\infty = v''v^\infty$. W.l.o.g. we assume that $|v'| > |v''|$. Since v' and v'' are suffixes of v, there exists a $\tilde{v} \in A^+$ such that $v' = \tilde{v}v''$. This implies $w^\infty = \tilde{v}v''v^\infty = \tilde{v}w^\infty$ and it follows that $w^\infty = \tilde{v}^\infty$. Note that $|\tilde{v}| < |v'| \leq |v|$. Thus we have found $u \in A^*$, $\tilde{v} \in A^+$ with $|\tilde{v}| < |v|$ and $v^\infty = u\tilde{v}^\infty$. This is a contradiction to the assumption of the lemma. ❑

The following lemma is an extension of Proposition 2.2. By this proposition, we find a permutation (induced by some word w) in every minimal dfa F where $L(F)$ is not starfree. Here we show that this permutation can be chosen in a minimal way, i.e., there do not exist words z, v with $|v| < |w|$ and $w^\infty = zv^\infty$. Note that this is nontrivial, since we have to prove that the existence of such words z, v indeed induces a permutation of *distinct* states.

Lemma 3.22. *Let $F = (A, S, \delta, s_0, S')$ be a minimal dfa.*

$L(F)$ is not starfree $\iff$ there exist $w \in A^+$, some $l \geq 2$ and distinct states $r_0, r_1, \dots r_{l-1} \in S$ such that

(i) there do not exist $z \in A^*$, $v \in A^+$ with $|v| < |w|$ and $w^\infty = zv^\infty$

(ii) $\delta(r_i, w) = r_{i+1}$ for $0 \leq i \leq l-1$ (with $r_l =_{\text{def}} r_0$)

Proof. "$\Longleftarrow$": This direction is an immediate consequence of Proposition 2.2.

"$\Longrightarrow$": Suppose that $L(F)$ is not starfree. Using Proposition 2.2 we choose a shortest $w \in A^+$, some $l \geq 2$ and distinct states $r_0, r_1, \dots, r_{l-1} \in S$ such that $\delta(r_i, w) = r_{i+1}$ for $0 \leq i \leq l-1$ (with $r_l =_{\text{def}} r_0$). In the following we will show that the violation of condition (i) implies that the choice of w was not minimal (which is a contradiction). This shows that w satisfies condition (i).

We choose a shortest word $v \in A^+$ and some $z \in A^*$, such that $|v| < |w|$ and $w^\infty = zv^\infty$. This implies that there do not exist $u' \in A^*$, $v' \in A^+$ with $|v'| < |v|$ and $v^\infty = u'v'^\infty$. Moreover, $w^\infty = zv^\infty$ implies the existence of some $u \in A^*$ such that $v^\infty = uw^\infty$ (simply delete the prefix z of w^∞). So we can apply Lemma 3.21 and obtain that $|w|$ is a multiple of $|v|$. This means that $|w| = n \cdot |v|$ and $w = v_1 v^{n-1} v_2$ for suitable $n \geq 2$ and $v_1, v_2 \in A^*$ with $v = v_2 v_1$.

Now we consider the sequence of states $(r'_i)_{i \geq 0}$ where $r'_i =_{\text{def}} \delta(r_0, v_1 v^i)$ for $i \geq 0$. From $w = v_1 v^{n-1} v_2$ and $v = v_2 v_1$ it follows that $\delta(r'_i, v^{l \cdot n - j}) = r'_0$ and $\delta(r'_i, v^{l \cdot n - j - 1} \cdot v_2) = r_0$ for all $i \geq 0$ and $j =_{\text{def}} (i \bmod l \cdot n)$. Suppose that there is some $i \geq 0$ with $r'_i = r'_{i+1}$. It follows that $\delta(r'_i, v) = r'_i$ and $\delta(r'_i, v_2) = r_0$. This implies $r_1 = \delta(r_0, w) = \delta(r'_i, v_2 w) = \delta(r'_i, v^n v_2) = r_0$, which is a contradiction to our assumption. It follows that $r'_i \neq r'_{i+1}$ for all $i \geq 0$.

Now choose a smallest j such that there is some $i < j$ with $r'_i = r'_j$ (such a j exists due to the finiteness of S). We have already seen that $j - i \geq 2$. Thus we have found a $v \in A^+$ and a list of $j - i \geq 2$ distinct states $r'_{i+1}, r'_{i+2}, \dots, r'_j$ such that $\delta(r'_{i'}, v) = r'_{i'+1}$ for $i + 1 \leq i' \leq j - 1$ (with $r'_j =_{\text{def}} r'_{i+1}$). Since $|v| < |w|$, this is a contradiction to the choice of the shortest $w \in A^+$ at the beginning of this proof. Hence our assumption was false, and we conclude that there do not exist $z \in A^*$, $v \in A^+$ with $|v| < |w|$ and $w^\infty = zv^\infty$. This proves our lemma. ❑

Lemma 3.23. *Let $\mathcal{T}$ be an initial pattern, $n \geq 1$, $l \geq 2$, $w \in A^+$ such that $\mathcal{C}_n^{\mathcal{T}} \subseteq SF$ and there do not exist $u \in A^*$, $v \in A^+$ with $|v| < |w|$ and $w^\infty = uv^\infty$. Then there exists a $p \in \mathbb{L}_n^{\mathcal{T}}$ such that for all $r \geq 3$ and $p' =_{\text{def}} \pi(p, r)$ the following holds:*

- *There exists some $u \in A^*$ with $w^\infty = u \cdot \left(\overline{p'}\right)^\infty$.*
- *The length of $\overline{p'}$ is a multiple of $|w|$, but it is not a multiple of $l \cdot |w|$.*

Proof. Choose alphabet letters a_i such that $w = a_1 \cdots a_m$. Let $F =_{\text{def}} (A, S, \delta, s_{1,1}, S')$ be the dfa where A is our fixed alphabet, $S =_{\text{def}} \left\{ s_{i,j} \mid 1 \leq i \leq l \text{ and } 1 \leq j \leq m \right\} \cup \{\tilde{s}\}$, $S' =_{\text{def}} S \setminus \{s_{1,1}\}$ and δ is defined as:

1. $\delta(s_{i,j}, a_j) = s_{i,j+1}$ for $1 \leq i \leq l$ and $1 \leq j < m$
2. $\delta(s_{i,m}, a_m) = s_{i+1,1}$ for $1 \leq i < l$

3. $\delta(s_{l,m}, a_m) = s_{1,1}$

4. $\delta(s, a) = \tilde{s}$ for all remaining arguments $(s, a) \in S \times A$

First of all let us determine the language accepted by F. For this we observe that $\tilde{s}$ is the only sink, and the initial state $s_{1,1}$ is the only rejecting state in the dfa F. Thus a word $v \in A^+$ is not in $L(F)$ if and only if $\delta(s_{1,1}, v) = s_{1,1}$. Moreover, the only possible path which starts at $s_{1,1}$ and which does not lead to the sink $\tilde{s}$ looks as follows:

$$s_{1,1} \xrightarrow{a_1} s_{1,2} \xrightarrow{a_2} \cdots \xrightarrow{a_m} s_{2,1} \xrightarrow{a_1} s_{2,2} \xrightarrow{a_2} \cdots \xrightarrow{a_m} s_{3,1} \xrightarrow{a_1} \quad \cdots \quad \xrightarrow{a_m} s_{l,1} \xrightarrow{a_1} s_{l,2} \xrightarrow{a_2} \cdots \xrightarrow{a_m} s_{1,1}$$

We go along this path if and only if the input is $(a_1 \cdots a_m)^l = w^l$. Therefore, a word $v \in A^+$ is not in $L(F)$ if and only if it is of the form w^{il}. This shows $L(F) = A^+ \setminus \{ w^{il} \mid i \geq 1 \}$.

Furthermore, F is a minimal dfa. Otherwise there would exist different states $s_1, s_2 \in S$ with $\delta(s_1, v) \in S' \iff \delta(s_2, v) \in S'$ for all $v \in A^*$. Note that both states have to be different from $\tilde{s}$, since $\tilde{s}$ is an accepting sink and from all other states a rejecting state is reachable. So it must be that $s_1 = s_{i,j}$ and $s_2 = s_{i',j'}$. The length of a shortest non–empty word leading from $s_{i,j}$ ($s_{i',j'}$, resp.) to a rejecting state is equal to the length of the shortest non–empty word which leads from $s_{i,j}$ ($s_{i',j'}$, resp.) to $s_{1,1}$, since $s_{1,1}$ is the only rejecting state. Observe that these lengths equal $m_1 =_{\text{def}} (m+1-j) + (l-i) \cdot |w|$ if we start at $s_{i,j}$, and $m_2 =_{\text{def}} (m+1-j') + (l-i') \cdot |w|$ if we start at $s_{i',j'}$. By assumption, these lengths have to be equal. So we obtain $j + i \cdot |w| = j' + i' \cdot |w|$. From $1 \leq j, j' \leq |w|$ it follows that $i = i'$ and $j = j'$ which is a contradiction to the choice of $s_{i,j}$ and $s_{i',j'}$. This shows that F is a minimal dfa.

So we have a minimal dfa F, a word $w \in A^+$ and states $s_{1,1}, s_{2,1}, \ldots, s_{l,1}$ such that $\delta(s_{i,1}, w) = s_{i+1,1}$ for $1 \leq i \leq l-1$ and $\delta(s_{l,1}, w) = s_{1,1}$. From Proposition 2.2 it follows that $L(F)$ is not starfree. From $L(F) \subseteq A^+$ and $\mathcal{C}_n^{\mathcal{T}} \subseteq SF$ we obtain that F has pattern $\mathbb{L}_n^{\mathcal{T}}$. By Definition 3.5, there exist $s_1, s_2 \in S$, $u, z \in A^*$, $p \in \mathbb{L}_n^{\mathcal{T}}$ such that $\delta(s_{1,1}, u) = s_1$, $\delta(s_1, z) \in S'$, $\delta(s_2, z) \notin S'$ and the states s_1, s_2 are connected via p. Note that s_1 and s_2 are different from $\tilde{s}$, since rejecting states are reachable from s_1 and s_2 (e.g. the state $\delta(s_1, \overline{p}z) = \delta(s_2, z)$ is rejecting). So we have $s_1 = s_{i_1,j_1}$ and $s_2 = s_{i_2,j_2}$ for suitable i_1, i_2, j_1, j_2.

We consider now an arbitrary state $s_{i,j} \in S$. By the construction of F, the following holds for all words $v \in A^*$:

$$\delta(s_{i,j}, v) \neq \tilde{s} \textit{ if and only if } v \textit{ is a prefix of } w_{i,j} =_{\text{def}} a_j a_{j+1} \cdots a_m \cdot w^{\infty}. \tag{17}$$

Let $r \geq 3$ and $p' =_{\text{def}} \pi(p, r)$. By Lemma 3.10, the states s_1, s_2 are also connected via p' and $p' \in \mathbb{L}_n^{\mathcal{T}}$. Moreover, by Lemma 3.6 we have $\overline{p'}, \overline{p'}^{\circ} \in A^+$ (for this it is necessary that $n \geq 1$), $\delta(s_1, \overline{p'}) = s_2$, $\delta(s_1, \overline{p'}^{\circ}) = s_1$ and $\delta(s_2, \overline{p'}^{\circ}) = s_2$. From (17) it follows that $\overline{p'}$ is a prefix of w_{i_1,j_1} and $w_{i_1,j_1} = \left(\overline{p'}^{\circ}\right)^{\infty} = w_{i_2,j_2}$. Furthermore, we have $w_{i_1,j_1} = \overline{p'} \cdot w_{i_2,j_2}$, since $\delta(s_1, \overline{p'}) = s_2$. This yields the equation $w_{i_1,j_1} = \overline{p'} \cdot w_{i_1,j_1}$, and we obtain $w_{i_1,j_1} = \left(\overline{p'}\right)^{\infty}$. From the definition of w_{i_1,j_1} it follows that $w^{\infty} = u \cdot w_{i_1,j_1}$ where $u =_{\text{def}} a_1 a_2 \cdots a_{j_1 - 1}$. This shows $w^{\infty} = u \cdot \left(\overline{p'}\right)^{\infty}$, and the first statement of the lemma follows.

From our assumption and Lemma 3.21 it follows that the length of $\overline{p'}$ is a multiple of $|w|$. Suppose that the length of $\overline{p'}$ is a multiple of $l \cdot |w|$. Then from the construction of F it follows that $s_2 = \delta(s_1, \overline{p'}) = s_1$. This is a contradiction to the choice of s_1, s_2 and z such that $\delta(s_1, z) \in S'$ and $\delta(s_2, z) \notin S'$. This shows the second statement of the lemma. ❑

Lemma 3.24. *Let $F = (A, S, \delta, s_0, S')$ be a minimal dfa, $\mathcal{T}$ an initial pattern and $n \geq 1$ such that $\mathcal{C}_n^{\mathcal{T}} \subseteq SF$. $L(F)$ is not starfree if and only if there exist a $p \in \mathbb{L}_n^{\mathcal{T}}$, some $l \geq 2$ and distinct states $r_0, r_1, \ldots r_{l-1} \in S$ such that r_i, r_{i+1} are connected via p for $0 \leq i \leq l-1$ (with $r_l =_{\text{def}} r_0$).*

Proof. The if–part follows from Lemma 2.2, since $\delta(r_i, w) = r_{i+1}$ for $w =_{\mathrm{def}} \overline{p}$ and $0 \le i \le l-1$ (Lemma 3.6.2).

Let $r =_{\mathrm{def}} |F|$ and assume that $L(F)$ is not starfree (if $|F| < 3$ then set $r =_{\mathrm{def}} 3$). By Lemma 3.22, there exist $w \in A^+$, some $l \ge 2$ and distinct states $r_0, r_1, \dots r_{l-1} \in S$ such that (i) there do not exist $u \in A^*$, $v \in A^+$ with $|v| < |w|$ and $w^\infty = uv^\infty$ and (ii) $\delta(r_i, w) = r_{i+1}$ for $0 \le i \le l-1$ (with $r_l =_{\mathrm{def}} r_0$). By Lemma 3.23, there exists a $p \in \mathbb{L}_n^{\mathcal{T}}$ such that for $p' =_{\mathrm{def}} \pi(p, r)$ it holds that (i) there exists a $u \in A^*$ with $w^\infty = u \cdot \left(\overline{p'}\right)^\infty$ and (ii) the length of $\overline{p'}$ is a multiple of $|w|$, but it is not a multiple of $l \cdot |w|$. Hence $|\overline{p'}| = m \cdot |w|$ and $\overline{p'} = w_1 w^{m-1} w_2$ for suitable $m \ge 1$, $w_1, w_2 \in A^*$ with $w = w_2 w_1$ (note that $m = 0$ is not possible, since $|\overline{p'}| > 0$ by Lemma 3.6.4).

Now we consider the sequence of states $(r'_i)_{i \ge 0}$ where $r'_i =_{\mathrm{def}} \delta(r_0, w_2 \cdot \left(\overline{p'}\right)^i)$ for $i \ge 0$. Suppose that there is some $i \ge 0$ with $r'_i = r'_{i+1}$. Note that

$$\begin{aligned}
\delta(r'_i, w_1) &= \delta(r_0, w_2 \cdot \left(\overline{p'}\right)^i w_1) \\
&= \delta(r_0, w_2 w_1 w^{m \cdot i - 1} w_2 w_1) \\
&= \delta(r_0, w^{m \cdot i + 1}) = r_{j'}
\end{aligned}$$

for $j' =_{\mathrm{def}} ((m \cdot i + 1) \mod l)$. It follows that

$$\begin{aligned}
\delta(r_{j'}, w^m) &= \delta(r'_i, w_1 w^{m-1} w_2 w_1) \\
&= \delta(r'_i, \overline{p'} w_1) = \delta(r'_{i+1}, w_1) \\
&= \delta(r'_i, w_1) = r_{j'}
\end{aligned}$$

By the choice of the states $r_0, \dots, r_{l-1}$ this implies that m is a multiple of l. From $|\overline{p'}| = m \cdot |w|$ it follows that $|\overline{p'}|$ is a multiple of $l \cdot |w|$. This is a contradiction to the choice of p. We conclude that $r'_i \ne r'_{i+1}$ for all $i \ge 0$.

From the sequence $(r'_i)_{i \ge 0}$ we choose an earliest r'_j such that there is some $r'_i = r'_j$ with $i < j$. We have already seen that $j - i \ge 2$. Thus we have found a list of $j - i \ge 2$ distinct states $r'_{i+1}, r'_{i+2}, \dots, r'_j$ such that $\delta(r'_{i'}, \overline{p'}) = r'_{i'+1}$ for $i + 1 \le i' \le j$ (with $r'_{j+1} = r'_{i+1}$). From Lemma 3.10.6 it follows that the states $r'_{i'}, r'_{i'+1}$ are connected via $p' \in \mathbb{L}_n^{\mathcal{T}}$ for $i + 1 \le i' \le j$. This proves the only–if–part of our lemma. ❑

Lemma 3.25. *If $\mathcal{C}_n^{\mathcal{T}} \subseteq SF$ for some initial pattern $\mathcal{T}$ and $n \ge 1$, then $\mathcal{C}_{n+1}^{\mathcal{T}} \subseteq SF$.*

Proof. Let $\mathcal{T}$ be an initial pattern and $n \ge 1$ such that $\mathcal{C}_n^{\mathcal{T}} \subseteq SF$. Moreover, let $F = (A, S, \delta, s_0, S')$ be a minimal dfa such that $L(F) \subseteq A^+$ is not starfree, we will show that $L(F) \notin \mathcal{C}_{n+1}^{\mathcal{T}}$. By Lemma 3.24, there exist a $p_0 \in \mathbb{L}_n^{\mathcal{T}}$, some $l \ge 2$ and distinct states $r_0, r_1, \dots r_{l-1} \in S$ such that r_i, r_{i+1} are connected via p_0 for $0 \le i \le l-1$ (with $r_l =_{\mathrm{def}} r_0$).

Now define $p =_{\mathrm{def}} \left(\overline{p_0}^{l-1}, p_0\right)$ and observe that $p \in \mathbb{L}_{n+1}^{\mathcal{T}}$. First of all we will show that the states r_0, r_{l-1} are connected via p. From Lemma 3.6.2 it follows that $\delta(r_i, \overline{p_0}) = r_{i+1}$ for $0 \le i \le l-1$. Hence $\delta(r_0, \overline{p_0}^{l-1}) = r_{l-1}$ and the states r_{l-1}, r_0 are connected via p_0. This shows that p appears at r_0, and analogously we obtain that p appears at r_{l-1}. Moreover, $\delta(r_0, \overline{p_0}^{l-1}) = r_{l-1}$ and p_0 appears at state r_{l-1} (Lemma 3.6.1). This shows that r_0, r_{l-1} are connected via p. Analogously one shows for $0 \le i \le l-1$ that the states r_{i+1}, r_i are connected via p.

Since F is minimal, there exist i, j with $0 \le i < j \le l-1$ and a word $z \in A^*$ such that $\delta(r_i, z) \in S' \iff \delta(r_j, z) \notin S'$. Since the states $r_0, \dots, r_{l-1}$ form a cycle, there exist also an i with $0 \le i \le l-1$ such that $\delta(r_{i+1}, z) \in S'$ and $\delta(r_i, z) \notin S'$. Furthermore, there exists a $u \in A^*$ with $\delta(s_0, u) = r_{i+1}$ (by the minimality of F). Since we have already seen that r_{i+1}, r_i are connected via p, it follows that F has pattern $\mathbb{L}_{n+1}^{\mathcal{T}}$. This shows $L(F) \notin \mathcal{C}_{n+1}^{\mathcal{T}}$, and it follows that $\mathcal{C}_{n+1}^{\mathcal{T}} \subseteq SF$. ❑

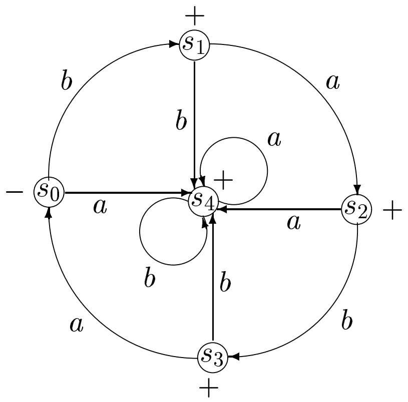
Theorem 3.26. *Let $\mathcal{T}$ be an initial pattern and $n \geq 1$.*

$$\bigcup_{i\geq 0} \mathcal{C}_i^{\mathcal{T}} \subseteq SF \quad \Longleftrightarrow \quad \mathcal{C}_n^{\mathcal{T}} \subseteq SF$$

Proof. Let $\mathcal{T}$ be an initial pattern and $n \geq 1$. It suffices to show that $\mathcal{C}_n^{\mathcal{T}} \subseteq SF$ implies $\mathcal{C}_i^{\mathcal{T}} \subseteq SF$ for all $i \geq 0$. By Lemma 3.25, this implication holds for all $i \geq n$.

If $\mathcal{C}_n^{\mathcal{T}} \subseteq SF$ then also $\mathrm{co}\mathcal{C}_n^{\mathcal{T}} \subseteq SF$, since SF is closed under complementation. From Lemma 3.19.2 it follows that $\mathcal{C}_{n-1}^{\mathcal{T}} \subseteq SF$. If we use this argument repeatedly, we obtain $\mathcal{C}_n^{\mathcal{T}} \subseteq SF \Longrightarrow \mathcal{C}_i^{\mathcal{T}} \subseteq SF$ for all $0 \leq i < n$. ❑

Remark 3.27. *It is necessary that we assume $n \geq 1$ in Theorem 3.26. In fact, this theorem does not hold for $n = 0$. For this end we consider a two letter alphabet $A = \{a, b\}$ and the initial pattern $\mathcal{T} =_{\mathrm{def}} \{(\varepsilon, \varepsilon), (\varepsilon, a), (\varepsilon, b)\}$. With the help of the known forbidden pattern for level $1/2$ of the Straubing–Thérien hierarchy (cf. Figure 4) we observe that $\mathcal{C}_0^{\mathcal{T}} = \mathcal{L}_{1/2} \subseteq SF$. In contrast, we will see that $\mathcal{C}_1^{\mathcal{T}} \not\subseteq SF$. To see this we have a look at the following dfa $F = (A, S, \delta, s_0, S')$.*



It is easy to see that F is a minimal dfa, and that it is not permutationfree. By Proposition 2.2, this implies $L(F) \notin SF$. It remains to show that $L(F) \in \mathcal{C}_1^{\mathcal{T}}$, i.e., we have to see that F does not have pattern $\mathbb{L}_1^{\mathcal{T}}$.

Assume that there are states s', s'', words u, z and a $p \in \mathbb{L}_1^{\mathcal{T}}$ such that $\delta(s_0, u) = s'$, $\delta(s', z) \in S'$, $\delta(s'', z) \notin S'$ and the states s', s'' are connected via p. Choose suitable $w_i \in A^+$ and $p_i \in \mathcal{T}$ such that $p = (w_0, p_0, \ldots, w_m, p_m)$. So we have $\overline{p} = w_0 \cdots w_m$ and $\overline{p}^\circ = w_0\overline{p_0} \cdots w_m\overline{p_m}$. Note that $s', s'' \in \{s_0, \ldots, s_3\}$, since rejecting states are reachable from s' and s''. It follows that $\overline{p}$ and $\overline{p}^\circ$ are alternating sequences of letters a, b, since all other words lead to the sink s_4.

Note that $\overline{p_i} \in \{\varepsilon, a, b\}$ for all $0 \leq i \leq m$, and assume that there exists a $0 \leq j < m$ with $\overline{p_j} \neq \varepsilon$. It follows that either $\overline{p_j}$ is equal to the last letter of w_j or it is equal to the first letter of w_{j+1}. This is a contradiction to the fact that $\overline{p}$ is an alternating sequences of letters a, b. It follows that $\overline{p}^\circ = \overline{p} \cdot \overline{p_m}$ where $\overline{p_m} \in \{a, b\}$, since $\overline{p} \neq \overline{p}^\circ$. W.l.o.g. we assume that $\overline{p_m} = a$. Hence $\delta(s'', a) = \delta(s', \overline{p}a) = s'$, and it follows that $\delta(s', a) = s_4$ and $\delta(s'', b) = s_4$. Therefore, at least one of the states s' and s'' does not have a $\overline{p}^\circ$-loop. This is a contradiction to our assumption, and it follows that F does not have pattern $\mathbb{L}_1^{\mathcal{T}}$. This shows $\mathcal{C}_1^{\mathcal{T}} \not\subseteq SF$.

Together we see that $\mathcal{C}_0^{\mathcal{T}} \subseteq SF$ and $\mathcal{C}_0^{\mathcal{T}} \subseteq \mathcal{C}_1^{\mathcal{T}} \not\subseteq SF$ since $\mathbb{L}_0^{\mathcal{T}} \preceq \mathbb{L}_1^{\mathcal{T}}$ which is easy to verify.

3.6 Decidability of the Pattern Classes

In this subsection we treat the decidability aspects of the pattern classes. It will turn out that $\mathbb{L}_n^{\mathcal{T}}$ is decidable in nondeterministic logarithmic space whenever a certain decision problem for the initial pattern $\mathcal{T}$

is decidable in these space bounds. Note that the decidability of the pattern classes has to depend on the initial pattern, since an undecidable set $\mathcal{T}$ (which can be easily constructed) leads to undecidable pattern classes.

We start with the definition of two problems addressing the question for the existence of paths (cf. Definition 3.28) and patterns (cf. Definition 3.29) which appear simultaneously in a dfa. In the Lemmas 3.30 and 3.31 we investigate the decidability of these problems, and at the end this leads to the desired results.

Definition 3.28. *Let $k \geq 1$. We define* REACH_k *to be the set of pairs (F, W) such that:*

1. *$F = (A, S, \delta, s_0, S')$ is a dfa*
2. *$W \subseteq S \times S$ with $|W| \leq k$*
3. *There exists a word $w \in A^+$ such that $\delta(s, w) = t$ for all $(s, t) \in W$.*

Definition 3.29. *Let $\mathcal{T}$ be an initial pattern, $n \geq 0$ and $k \geq 1$. We define* $\mathrm{PATTERN}^{\mathcal{T}}_{n,k}$ *to be the set of all triples (F, T_1, T_2) such that the following holds:*

1. *$F = (A, S, \delta, s_0, S')$ is a dfa*
2. *$T_1 \subseteq S$ with $|T_1| \leq k$*
3. *$T_2 \subseteq S \times S$ with $|T_2| \leq k$*
4. *There exists a $p \in \mathbb{L}^{\mathcal{T}}_n$ such that for all $s \in T_1$ and all $(q, r) \in T_2$ it holds that (i) p appears at s and (ii) q, r are connected via p.*

We want to make precise how we think of a dfa as an input to a Turing machine. Therefore, we give an explicit encoding as follows. Using the three–letter alphabet $\{0, |, \#\}$ we want to encode a dfa $F = (A, S, \delta, s_0, S')$. For this we fix arbitrary orderings on the sets A, S and S', such that we obtain $A = \{a_1, \ldots, a_{|A|}\}$, $S = \{s_1, \ldots, s_{|S|}\}$ (one of them is the starting state s_0) and $S' = \{s'_1, \ldots, s'_{|S'|}\}$. Moreover, we identify the elements a_i, s_j of the sets A, S with their index numbers i, j (resp.). Now we can encode F in the following way.

$$\underbrace{0^{|A|}}_{A}\#\underbrace{0^{|S|}}_{S}\#\underbrace{0^{\delta(s_1,a_1)}|0^{\delta(s_1,a_2)}|\cdots|0^{\delta(s_1,a_{|A|})}|0^{\delta(s_2,a_1)}|0^{\delta(s_2,a_2)}|\cdots|0^{\delta(s_{|S|},a_{|A|})}}_{\delta}\#\underbrace{0^{s_0}}_{s_0}\#\underbrace{0^{s'_1}|\cdots|0^{s'_{|S'|}}}_{S'}$$

The sets W, S_1, S_2 in the Definitions 3.28 and 3.29 are encoded analogous to S' (we encode a pair $(q, r) \in S \times S$ by $0^{s_1 \cdot |S| + s_2}$). Taking the respective codes together (separated by $\#$ signs), we obtain codes for (F, W) and (F, S_1, S_2). It is easy to see that on input of a word from $\{0, |, \#\}^*$, we can check in deterministic logarithmic space whether this word is a valid representation of a pair (F, W) (of a triple (F, S_1, S_2), resp.). Therefore, in forthcoming investigation of algorithms we may assume that all inputs are valid representations.

Lemma 3.30. *For $k \geq 1$ we have* $\mathrm{REACH}_k \in \mathrm{NL}$.

Proof. We use a slight modification of the algorithm solving the graph accessibility problem. If $|W| = 0$ then we are done. Otherwise we assign the elements of W to program variables $s_1, \ldots, s_k$ and $t_1, \ldots, t_k$ (some may take the same value if $|W| < k$). Now we guess a word $w \in A^+$ letter by letter, and we simultaneous follow the paths which start at $s_1, \ldots, s_k$ and which are labeled with w. Moreover, in each step we guess whether we have already reached the end of w, and if so, we check whether $s_i = t_i$ for all $1 \leq i \leq k$. ❑

We will consider oracle machines working in nondeterministic logarithmic space which have the following access to the oracle. The machine has an (read–only) input tape, a (write–only) query tape and a (read-write) working tape which is bounded logarithmically in the input size. Furthermore, from the moment where we write the first letter on the query tape, we are not allowed to make nondeterministic branches until we ask the oracle. After doing this we obtain the corresponding answer and the query tape is empty. Using this model, introduced in [RST82], we can prove the following lemma. We assume that the machine represents a single state of a dfa on its working tape in binary by the index number of the state. Hence the space needed to do this is bounded logarithmically in the input size.

Lemma 3.31. *Let $\mathcal{T}$ be an initial pattern, then* $\textsc{Pattern}^{\mathcal{T}}_{n,k} \in \mathrm{NL}^{\textsc{Pattern}^{\mathcal{T}}_{(n-1),3k}}$ *for each $n \geq 0$ and each $k \geq 1$.*

Proof. In Table 1 we describe a nondeterministic algorithm having access to a $\textsc{Reach}_{4k}$ oracle and to a $\textsc{Pattern}^{\mathcal{T}}_{(n-1),3k}$ oracle. The notations in this table are adopted from the Figures 1 and 2. We will show that this algorithm works in logarithmic space and decides $\textsc{Pattern}^{\mathcal{T}}_{n,k}$. By Lemma 3.30 we have $\textsc{Reach}_{k} \in \mathrm{NL}$. Since the access to an NL oracle does not rise the power of an NL machine, i.e., $\mathrm{NL}^{\mathrm{NL}} = \mathrm{NL}$ [RST82, Sze87, Imm88], we can go without the $\textsc{Reach}_{4k}$ oracle and obtain the desired algorithm.

First of all we want to observe that the algorithm accesses the oracle in the way as described above. For this we only have to consider step 4. Since on the one hand we have already computed the sets W, T'_1 and T'_2 (they are stored on the working–tape) and on the other hand F is stored on the input–tape, we can actually write down the queries (F, W) and (F, T'_1, T'_2) without making any nondeterministic branches.

Let us analyse the space on the working–tape which is needed on input (F, T_1, T_2). Note that our program uses only a constant number of variables (this number can be bounded by a function of $O(k)$, and k is a constant). Moreover, all variables except T'_1, T'_2, W contain index numbers of states of F, which can be stored in logarithmic space. Each of the variables T'_1, T'_2, W contains a set consisting of at most $4k$ (pairs of) index numbers of states. Note also that we can produce the encoding of the queries as needed for the oracle input deterministically with a logarithmic space bound on the working tape. This shows that our algorithm works in logarithmic space.

In the remaining part of this proof we will show that our algorithm decides $\textsc{Pattern}^{\mathcal{T}}_{n,k}$. First of all we want to see that the computation has an accepting path if $(F, T_1, T_2) \in \textsc{Pattern}^{\mathcal{T}}_{n,k}$. For this let $p = (w_0, p_0, \ldots, w_m, p_m) \in \mathbb{L}^{\mathcal{T}}_n$ be a witnessing pattern (cf. Definition 3.29.4). We denote the involved states of the appearance of p at s_i as in Figure 1, and we denote the involved states of the connection of q_j, r_j via p as in Figure 2. Now consider that path of the computation where we carry out exactly $m + 1$ passes of the loop and where we guess the states $\phi_{i,l}, \psi_{i,l}, \alpha_{j,l}, \beta_{j,l}, \gamma_{j,l}, \delta_{j,l}, \lambda_{j,l}$ at the beginning of the l-th pass of the loop (starting with pass 0). It can be easily verified that this is an accepting path.

Now suppose that the computation on input (F, T_1, T_2) has an accepting path, and fix one of these paths. Choose m such that on this path the loop is passed $m + 1$ times. Note that in each pass of the loop we receive positive answers to the queries $(F, W) \in \textsc{Reach}_{4k}$ and $(F, T'_1, T'_2) \in \textsc{Pattern}^{\mathcal{T}}_{(n-1),3k}$ (otherwise the fixed path would be rejecting). It follows that for each pass l there exists a word $w_l \in A^+$ witnessing $(F, W) \in \textsc{Reach}_{4k}$, and there exists a pattern $p_l \in \mathbb{L}^{\mathcal{T}}_{n-1}$ witnessing $(F, T'_1, T'_2) \in \textsc{Pattern}^{\mathcal{T}}_{(n-1),3k}$. Now define $p =_{\mathrm{def}} (w_0, p_0, \ldots, w_m, p_m)$. Using the states $\phi_{i,l}, \psi_{i,l}, \alpha_{j,l}, \beta_{j,l}, \gamma_{j,l}, \delta_{j,l}, \lambda_{j,l}$ which were guessed at the beginning of the l-th pass of the loop, we can verify that (i) p appears at all $s_i \in T_1$ and (ii) all q_i, r_i with $(q_i, r_i) \in T_2$ are connected via p. ❑

Corollary 3.32. *Let $\mathcal{T}$ be an initial pattern such that* $\textsc{Pattern}^{\mathcal{T}}_{0,k} \in \mathrm{NL}$ *for all $k \geq 1$. Then* $\textsc{Pattern}^{\mathcal{T}}_{n,k} \in \mathrm{NL}$ *for each $n \geq 0$ and each $k \geq 1$.*

Proof. We prove this by induction on n. The induction base is by assumption. The induction step follows from Lemma 3.31 and the fact that $\mathrm{NL}^{\mathrm{NL}} = \mathrm{NL}$ [RST82, Sze87, Imm88]. ❑

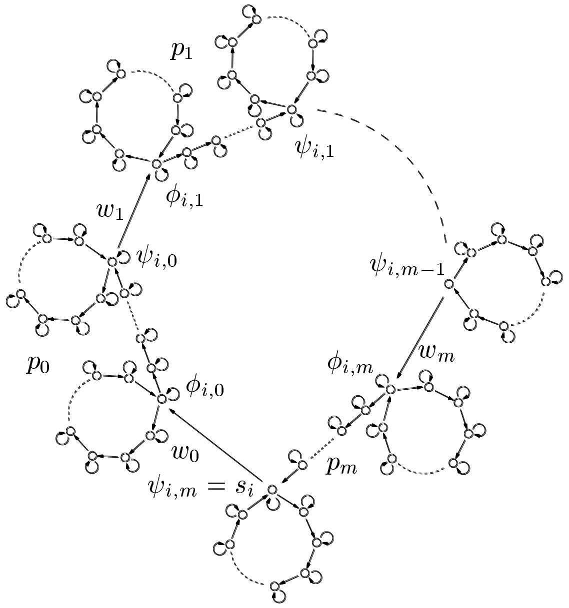


Figure 1: Example for an appearance of a pattern at some state s_i.

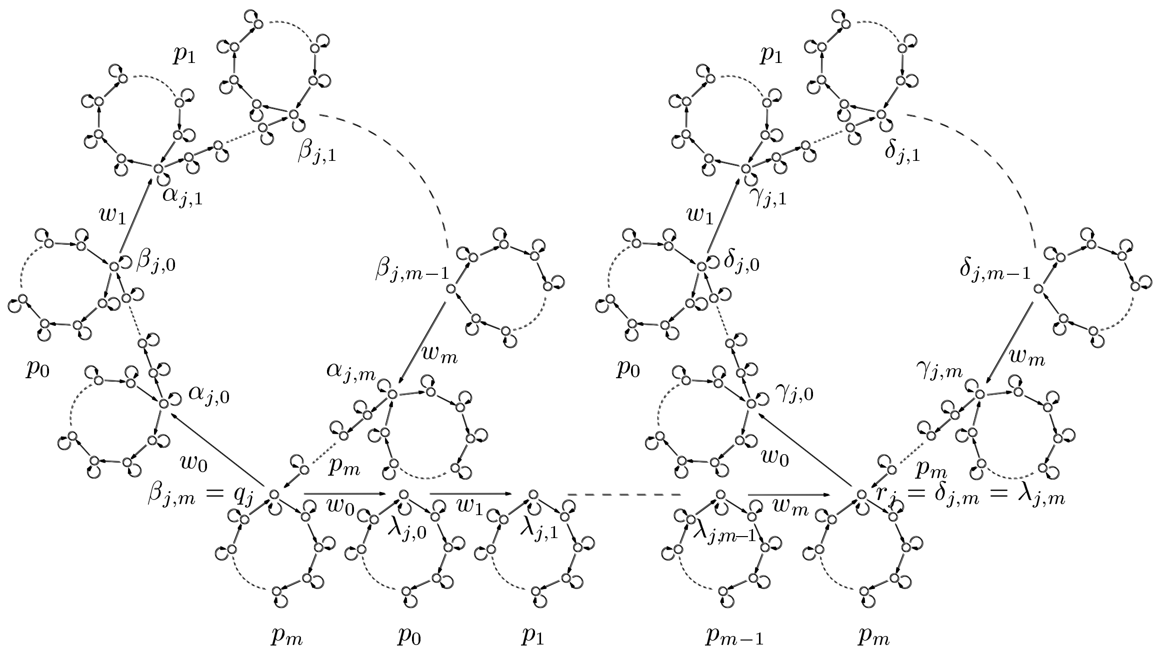


Figure 2: Example for a connection of two states q_j, r_j via a pattern.

Step, Label	Command	Remark
1.	Let $t_1 := \lvert T_1\rvert$, $t_2 := \lvert T_2\rvert$ and let s_i, q_j, r_j such that $T_1 = \{s_1, \ldots, s_{t_1}\}$ and $T_2 = \{(q_1, r_1), \ldots, (q_{t_2}, r_{t_2})\}$.	Note that t_1, t_2 are bounded by the constant k. We have to decide whether there is a $p \in \mathbb{L}_n^{\mathcal{T}}$ such that (i) p appears at all s_i (Figure 1) and (ii) all q_j, r_j are connected via p (Figure 2).
2.	For $1 \leq i \leq t_1$ and $1 \leq j \leq t_2$ let: $\begin{array}{ll} \psi_i^{\text{start}} := s_i & \beta_j^{\text{start}} := q_j \\ \delta_j^{\text{start}} := r_j & \lambda_j^{\text{start}} := q_j \end{array}$	Variables marked with 'start' contains the starting point from where we have to guess and check the next fragment of the pattern.
3. `loop:`	Guess states ϕ_i, ψ_i for $1 \leq i \leq t_1$, states $\alpha_j, \beta_j, \gamma_j, \delta_j, \lambda_j$ for $1 \leq j \leq t_2$ and let: $\begin{array}{lcl} T_1' & := & \{\, \lambda_j \mid 1 \leq j \leq t_2 \,\} \\ T_2' & := & \{\, (\phi_i, \psi_i) \mid 1 \leq i \leq t_1 \,\} \cup \\ & & \{\, (\alpha_j, \beta_j) \mid 1 \leq j \leq t_2 \,\} \cup \\ & & \{\, (\gamma_j, \delta_j,) \mid 1 \leq j \leq t_2 \,\} \\ W & := & \{\, (\psi_i^{\text{start}}, \phi_i) \mid 1 \leq i \leq t_1 \,\} \cup \\ & & \{\, (\beta_j^{\text{start}}, \alpha_j) \mid 1 \leq j \leq t_2 \,\} \cup \\ & & \{\, (\delta_j^{\text{start}}, \gamma_j) \mid 1 \leq j \leq t_2 \,\} \cup \\ & & \{\, (\lambda_j^{\text{start}}, \lambda_j) \mid 1 \leq j \leq t_2 \,\} \end{array}$	The guessed states correspond to Figures 1 and 2. In the l-th pass of this loop (starting with pass 0) the variables $\phi_i, \psi_i, \alpha_j, \beta_j, \gamma_j, \delta_j, \lambda_j$ correspond to $\phi_{i,l}, \psi_{i,l}, \alpha_{j,l}, \beta_{j,l}, \gamma_{j,l}, \delta_{j,l}, \lambda_{j,l}$ (respectively). Moreover, at the beginning of the l-th pass we have a correspondence between $\psi_i^{\text{start}}, \beta_j^{\text{start}}, \delta_j^{\text{start}}, \lambda_j^{\text{start}}$ and $\psi_{i,l-1}, \beta_{j,l-1}, \delta_{j,l-1}, \lambda_{j,l-1}$ (resp.). Using T_1' and T_2' we will ask the oracle whether there is a pattern p_l which connects (and appears at) the guessed states. With W we will test the existence of a word w_l (cf. Figures 1 and 2).
4.	Ask the following queries and reject when a negative answer is given. $(F, W) \in \textsc{Reach}_{4k}$ $(F, T_1', T_2') \in \textsc{Pattern}_{(n-1),3k}^{\mathcal{T}}$	If at least one negative answer is given then the states guessed at the previous step do not correspond to a pattern from $\mathbb{L}_n^{\mathcal{T}}$.
5.	For $1 \leq i \leq t_1$ and $1 \leq j \leq t_2$ let: $\begin{array}{ll} \psi_i^{\text{start}} := \psi_i & \beta_j^{\text{start}} := \beta_j \\ \delta_j^{\text{start}} := \delta_j & \lambda_j^{\text{start}} := \lambda_j \end{array}$	Here we set the next starting points.
6.	Jump nondeterministically to `loop` or to `exit`.	Guess whether we have already checked the right number of fragments of the pattern, i.e., whether the number of passes equals m.
7. `exit:`	Accept if and only if the following conditions hold for all $1 \leq i \leq t_1$ and $1 \leq j \leq t_2$: $\begin{array}{ll} \psi_i = s_i & \beta_j = q_j \\ \delta_j = r_j & \lambda_j = r_j \end{array}$	It remains to check whether the guessed loops have reached their starting points, and whether the path which was guessed via λ_i leads from q_i to r_i.

Table 1: An algorithm which decides $(F, T_1, T_2) \in \textsc{Pattern}_{n,k}^{\mathcal{T}}$ on input of a dfa $F = (A, S, \delta, s_0, S')$ and sets $T_1 \subseteq S$ and $T_2 \subseteq S \times S$ with $\lvert T_1\rvert, \lvert T_2\rvert \leq k$.

Corollary 3.33. *Let $\mathcal{T}$ be an initial pattern such that* $\text{PATTERN}^{\mathcal{T}}_{0,k} \in \text{NL}$ *for each* $k \geq 1$*. Then for a fixed* $n \geq 0$ *it is decidable in nondeterministic logspace whether a given dfa* F *has pattern* $\mathbb{L}^{\mathcal{T}}_n$*.*

Proof. On input $F = (A, S, \delta, s_0, S')$ we guess states $s_1, s_2, s^+, s^- \in S$ and check whether $s^+ \in S'$ and $s^- \notin S'$. Now we test $(F, \{(s_0, s_1)\}) \in \text{REACH}_1$ and $(F, \{(s_1, s^+), (s_2, s^-)\}) \in \text{REACH}_2$ which is possible in nondeterministic logspace by Lemma 3.30. It remains to check whether or not $(F, \emptyset, \{(s_1, s_2)\}) \in \text{PATTERN}^{\mathcal{T}}_{n,1}$ which is also possible in NL by Corollary 3.32. ❑

4 Consequences for Concatenation Hierarchies

From now on we consider two special initial patterns. We will see that the classes of languages being defined by these patterns are closely related to the dot–depth hierarchy and the Straubing–Thérien hierarchy.

Definition 4.1. *We define the following initial patterns.*

$$\begin{aligned}\mathcal{L} &=_{\text{def}} \{\varepsilon\} \times A^* \\ \mathcal{B} &=_{\text{def}} A^+ \times A^+\end{aligned}$$

It is easy to see that $\mathcal{L}$ and $\mathcal{B}$ are indeed initial patterns.

4.1 Some Auxiliary Results

We start with some easy to see results that make handling the lower levels of the pattern classes easier.

Proposition 4.2. *Let $p \in \mathbb{L}_1^{\mathcal{L}}$. Let $F = (A, S, \delta, s_0, S')$ be a dfa and $s \in S$. Then p appears at s if and only if $\delta(s, \overline{p}^\circ) = s$.*

Proof. Let $p = (w_0, p_0, \dots, w_m, p_m) \in \mathbb{L}_1^{\mathcal{L}}$ with $m \geq 0$ and $w_i \in A^+$ and $p_i = (l_i, b_i) \in \mathbb{L}_0^{\mathcal{L}} = \{\varepsilon\} \times A^*$ for all $0 \leq i \leq m$. First assume p appears at s. Then $\delta(s, \overline{p}^\circ) = s$ by Lemma 3.6.3. Conversely, let $\delta(s, \overline{p}^\circ) = s$ and define the witnessing states as $r_0 =_{\text{def}} \delta(s, w_0)$, $g_j =_{\text{def}} \delta(r_j, b_j)$ and $r_{j+1} =_{\text{def}} \delta(g_j, w_{j+1})$ for $0 \leq j < m$, and $g_m =_{\text{def}} \delta(r_m, b_m)$. One easily verifies with help of these states that p appears at s. Just observe that two states $r, g \in S$ are connected via some pattern $p' = (l', b') \in \mathbb{L}_0^{\mathcal{L}}$ if and only if $\delta(r, b') = g$. ❑

Lemma 4.3. *Let $p = (l, b) \in \mathbb{L}_0^{\mathcal{B}} = A^+ \times A^+$ such that l and b have the same first letter and $\alpha(b) \subseteq \alpha(l)$. Then there is some $p' \in \mathbb{L}_1^{\mathcal{L}}$ such that for all dfa $F = (A, S, \delta, s_0, S')$ and $s_1, s_2 \in S$ the following holds. If s_1 and s_2 are connected via p, then s_1 and s_2 are connected via p' and $\overline{p'}^\circ = l^n$ for $n =_{\text{def}} |b|$.*

Proof. Let $b = a_0 a_1 \cdots a_m$ for some $m \geq 0$ and $a_i \in A$. Because $\{a_0, \dots, a_m\} \subseteq \alpha(l)$ we can write l for all $0 \leq i \leq m$ as $l = l_i a_i l_i'$. We can assume that $l_0 = \varepsilon$. Now define $p' =_{\text{def}} (w_0', p_0', w_1', p_1', \dots, w_m', p_m')$ with $w_i' =_{\text{def}} a_i$ for $0 \leq i \leq m$, and $p_i' =_{\text{def}} (\varepsilon, l_j' l_{j+1})$ for $0 \leq j < m$ and $p_m' =_{\text{def}} (\varepsilon, l_m')$. By Definition 3.1 we have $p_i' \in \mathbb{L}_0^{\mathcal{L}}$ and with Definition 3.2 we see that $p' \in \mathbb{L}_1^{\mathcal{L}}$.

Observe with Definition 3.4 that $\overline{p'}^\circ = a_0 l_0' \cdot l_1 a_1 l_1' \cdots l_m a_m l_m' = l_0 a_0 l_0' \cdot l_1 a_1 l_1' \cdots l_m a_m l_m' = l^{m+1}$ and that $m + 1 = |b|$. It remains to show for some dfa $F = (A, S, \delta, s_0, S')$ and $s_1, s_2 \in S$ that s_1 and s_2 are connected via p' under the assumption that s_1 and s_2 are connected via p. From the latter we have by Definition 3.1 that $\delta(s_1, l) = s_1$ and $\delta(s_1, b) = s_2 = \delta(s_2, l)$. Hence, $\delta(s_1, l^{m+1}) = \delta(s_1, \overline{p'}^\circ) = s_1$ and $\delta(s_2, l^{m+1}) = \delta(s_2, \overline{p'}^\circ) = s_2$, and by Proposition 4.2 we see that p' appears at s_1 and at s_2. With the definition of witnessing states $\tilde{s}_0 =_{\text{def}} \delta(s_1, w_0')$, $\tilde{s}_{j+1} =_{\text{def}} \delta(\tilde{s}_j, w_{j+1}')$ for $0 \leq j < m$ and $\tilde{s}_m =_{\text{def}} s_2$ we get from Definition 3.4 and $\delta(s_1, b) = s_2$ that s_1 and s_2 are connected via p'. Note that p_i' appears at every state $s \in S$. ❑

4.2 Inclusion Relations and Lower Levels

Lemma 4.4. *The following holds.*

1. $\mathbb{L}_0^{\mathcal{L}} \preceq \mathbb{L}_0^{\mathcal{B}}$
2. $\mathbb{L}_0^{\mathcal{B}} \preceq \mathbb{L}_1^{\mathcal{L}}$
3. $\mathbb{L}_0^{\mathcal{L}} \preceq \mathbb{L}_1^{\mathcal{L}}$

4. $\mathbb{L}_0^{\mathcal{B}} \preceq \mathbb{L}_1^{\mathcal{B}}$

Proof. To see the first statement, let $p_1 = (v, w) \in \mathbb{L}_0^{\mathcal{B}} = A^+ \times A^+$ and define $p_2 = (\varepsilon, w)$. Obviously, $p_2 \in \mathbb{L}_0^{\mathcal{L}}$. Let $F = (A, S, \delta, s_0, S')$ be a dfa and $s, s_1, s_2 \in S$. Clearly, $\delta(s, \varepsilon) = s$, so p_2 appears at s. Now suppose that s_1, s_2 are connected via p_1. In particular $\delta(s_1, w) = s_2$, so s_1, s_2 are connected via p_2. This shows $\mathbb{L}_0^{\mathcal{L}} \preceq \mathbb{L}_0^{\mathcal{B}}$.

For the second statement let $p_1 = (w_0, p'_0, \ldots, w_m, p'_m) \in \mathbb{L}_1^{\mathcal{L}}$ with $m \geq 0$ and $w_i \in A^+$ and $p'_i = (l_i, b_i) \in \mathbb{L}_0^{\mathcal{L}} = \{\varepsilon\} \times A^*$ for all $0 \leq i \leq m$. Define $p_2 =_{\mathrm{def}} (\overline{p_1}^{\circ}, \overline{p_1})$. By Lemma 3.6.4 we have $p_2 \in \mathbb{L}_0^{\mathcal{B}} = A^+ \times A^+$. Again, let $F = (A, S, \delta, s_0, S')$ be a dfa and $s, s_1, s_2 \in S$. First assume that p_1 appears at s. By Lemma 3.6.3 we have $\delta(s, \overline{p_1}^{\circ}) = s$, and hence also p_2 appears at s. Now suppose s_1, s_2 are connected via p_1. Then p_1 appears at s_1 and at s_2, and so does p_2. Furthermore, $\delta(s_1, \overline{p_1}) = s_2$ by Lemma 3.6.2, so s_1, s_2 are connected via p_2. This shows $\mathbb{L}_0^{\mathcal{B}} \preceq \mathbb{L}_1^{\mathcal{L}}$.

Taking together the first and second statement, we get statement 3. The last statement has the same proof as the second one. ❑

Theorem 4.5. *For $n \geq 0$ the following holds.*

1. $\mathcal{C}_n^{\mathcal{L}} \subseteq \mathcal{C}_n^{\mathcal{B}}$
2. $\mathcal{C}_n^{\mathcal{B}} \subseteq \mathcal{C}_{n+1}^{\mathcal{L}}$
3. $\mathcal{C}_n^{\mathcal{L}} \cup \mathrm{co}\mathcal{C}_n^{\mathcal{L}} \subseteq \mathcal{C}_{n+1}^{\mathcal{L}} \cap \mathrm{co}\mathcal{C}_{n+1}^{\mathcal{L}}$
4. $\mathcal{C}_n^{\mathcal{B}} \cup \mathrm{co}\mathcal{C}_n^{\mathcal{B}} \subseteq \mathcal{C}_{n+1}^{\mathcal{B}} \cap \mathrm{co}\mathcal{C}_{n+1}^{\mathcal{B}}$

Proof. We have $\mathbb{L}_0^{\mathcal{L}} \preceq \mathbb{L}_0^{\mathcal{B}}$ and $\mathbb{L}_0^{\mathcal{B}} \preceq \mathbb{L}_1^{\mathcal{L}}$ by Lemma 4.4. By Lemma 3.17 this implies $\mathbb{L}_n^{\mathcal{L}} \preceq \mathbb{L}_n^{\mathcal{B}}$ and $\mathbb{L}_n^{\mathcal{B}} \preceq \mathbb{L}_{n+1}^{\mathcal{L}}$ for all $n \geq 0$. Now the first two statements follow from Lemma 3.18. From Lemma 4.4 we also know $\mathbb{L}_0^{\mathcal{L}} \preceq \mathbb{L}_1^{\mathcal{L}}$ and $\mathbb{L}_0^{\mathcal{B}} \preceq \mathbb{L}_1^{\mathcal{B}}$. Just apply Theorem 3.20 to get the remaining two statements. ❑

The reader may compare these inclusion relations with Propositions 2.7 and 2.8. However, the connections between pattern classes and classes of concatenation hierarchies are even closer (which explains the naming of the initial patterns).

Theorem 4.6. *For $n \geq 0$ the following holds.*

1. $\mathcal{L}_{1/2} = \mathcal{C}_0^{\mathcal{L}}$
2. $\mathcal{L}_{3/2} = \mathcal{C}_1^{\mathcal{L}}$
3. $\mathcal{L}_{n+1/2} \subseteq \mathcal{C}_n^{\mathcal{L}}$
4. $\mathcal{B}_{1/2} = \mathcal{C}_0^{\mathcal{B}}$
5. $\mathcal{B}_{3/2} = \mathcal{C}_1^{\mathcal{B}}$
6. $\mathcal{B}_{n+1/2} \subseteq \mathcal{C}_n^{\mathcal{B}}$

Proof. To show statements 1, 2 and 4 we recall forbidden pattern characterizations from [PW97]. For statement 5 we recall such a result from [GS00].

Statement 1: The following is shown in [PW97].

(a) Let F be a minimal dfa with $L(F) \subseteq A^*$. Then $L(F) \in \mathcal{L}_{1/2} \cup \{A^*\}$ if and only if F does not have a subgraph in its transition graph as depicted in Figure 4 with $u, w, z \in A^*$ and s^+ (s^-) being an accepting (rejecting) state.

This is [PW97, Theorem 8.5] after rewriting their notations, together with Theorem 2.20.

Now suppose $L \in \mathcal{L}_{1/2}$ and let $F = (A, S, \delta, s_0, S')$ be the minimal dfa with $L(F) = L$. Assume that F has pattern $\mathbb{L}_0^{\mathcal{L}}$, hence by Definition 3.5 there are $s_1', s_2' \in S$, $u', z' \in A^*$, $(\varepsilon, w') \in \mathbb{L}_0^{\mathcal{L}}$ such that $\delta(s_0, u') = s_1'$, $\delta(s_1', z') \in S'$, $\delta(s_2', z') \notin S'$ and $\delta(s_1', w') = s_2'$. But this is a subgraph exactly as stated in (a) above, contradicting $L \in \mathcal{L}_{1/2}$. Hence F does not have pattern $\mathbb{L}_0^{\mathcal{L}}$, so $L \in \mathcal{C}_0^{\mathcal{L}}$.

Conversely, let $L \in \mathcal{C}_0^{\mathcal{L}}$ and let $F = (A, S, \delta, s_0, S')$ be the minimal dfa with $L(F) = L$. Assume that F has a subgraph in its transition graph as stated in (a). Then $p =_{\text{def}} (\varepsilon, w) \in \mathbb{L}_0^{\mathcal{L}}$ and we see with Definition 3.5 that F has pattern $\mathbb{L}_0^{\mathcal{L}}$ since s_1, s_2 are connected via p, contradicting $L \in \mathcal{C}_0^{\mathcal{L}}$. So $L \in \mathcal{L}_{1/2}$ or $L = A^*$ by (a). The latter is not possible since $L \subseteq A^+$ by definition of $\mathcal{C}_0^{\mathcal{L}}$.

Statement 4: The following is shown in [PW97].

(b) Let F be a minimal dfa with $L(F) \subseteq A^+$. Then $L(F) \in \mathcal{B}_{1/2}$ if and only if F does not have a subgraph in its transition graph as depicted in Figure 7 with $u \in A^*$, $v, w, z \in A^+$, and s^+ (s^-) being an accepting (rejecting) state.

This is [PW97, Theorem 8.15] after rewriting their notations (recall also by Theorem 2.20 that we talk about the same class of languages $\mathcal{B}_{1/2}$).

Suppose $L \in \mathcal{B}_{1/2}$ and let $F = (A, S, \delta, s_0, S')$ be the minimal dfa with $L(F) = L$. Assume that F has pattern $\mathbb{L}_0^{\mathcal{B}}$, hence by Definition 3.5 there are $s_1', s_2' \in S$, $u', z' \in A^*$, $(v', w') \in \mathbb{L}_0^{\mathcal{B}} = A^+ \times A^+$ such that $\delta(s_0, u') = s_1' = \delta(s_1', v')$, $\delta(s_1', z') \in S'$, $\delta(s_2', z') \notin S'$ and $\delta(s_1', w') = s_2' = \delta(s_2', v')$. But this is a subgraph exactly as stated in (b) above, with $u =_{\text{def}} u'$, $v =_{\text{def}} v'$, $w =_{\text{def}} w'$ and $z =_{\text{def}} v'z'$, contradicting $L \in \mathcal{B}_{1/2}$. Hence F does not have pattern $\mathbb{L}_0^{\mathcal{B}}$, so $L \in \mathcal{C}_0^{\mathcal{B}}$.

Conversely, let $L \in \mathcal{C}_0^{\mathcal{B}}$ and let $F = (A, S, \delta, s_0, S')$ be the minimal dfa with $L(F) = L$. Assume that F has a subgraph in its transition graph as stated in (b). Then $p =_{\text{def}} (v, w) \in \mathbb{L}_0^{\mathcal{B}}$ and we see with Definition 3.5 that F has pattern $\mathbb{L}_0^{\mathcal{B}}$ since s_1, s_2 are connected via p, contradicting $L \in \mathcal{C}_0^{\mathcal{L}}$. So $L \in \mathcal{B}_{1/2}$.

Statement 2: The following is shown in [PW97].

(c) Let F be a minimal dfa with $L(F) \subseteq A^*$. Then $L(F) \in \mathcal{L}_{3/2} \cup \{L \cup \{\varepsilon\} : L \in \mathcal{L}_{3/2}\}$ if and only if F does not have a subgraph in its transition graph as depicted in Figure 7 with $u, v, w, z \in A^*$, $\alpha(v) = \alpha(w)$ and s^+ (s^-) being an accepting (rejecting) state.

This is [PW97, Theorem 8.9] after rewriting their notations, together with Theorem 2.20.

Suppose $L \in \mathcal{L}_{3/2}$ and let $F = (A, S, \delta, s_0, S')$ be the minimal dfa with $L(F) = L$. Assume that F has pattern $\mathbb{L}_1^{\mathcal{L}}$ and let $p = (w_0, p_0, \dots, w_m, p_m) \in \mathbb{L}_1^{\mathcal{L}}$ with $m \geq 0$ and $w_i \in A^+$ and $p_i = (l_i, b_i) \in \mathbb{L}_0^{\mathcal{L}} = \{\varepsilon\} \times A^*$ for all $0 \leq i \leq m$. By Definition 3.5 there are $s_1', s_2' \in S$ and $u', z' \in A^*$ such that $\delta(s_0, u') = s_1'$, $\delta(s_1', z') \in S'$, $\delta(s_2', z') \notin S'$ and s_1', s_2' are connected via p. Observe that $\overline{p}^\circ = w_0 b_0 w_1 b_1 \cdots w_m b_m \in A^+$ and $\overline{p} = w_0 w_1 \cdots w_m \in A^+$ by Lemma 3.6.4. Since p appears at s_1' we get from Lemma 3.6.3 that $\delta(s_1', \overline{p}^\circ) = s_1'$. The same holds for s_2'. Moreover, it holds that $\delta(s_1', \overline{p}) = s_2'$ by Lemma 3.6.2. So with $u =_{\text{def}} u'$, $v =_{\text{def}} \overline{p}^\circ$, $w =_{\text{def}} \overline{p}^\circ \overline{p}\, \overline{p}^\circ$ and $z =_{\text{def}} z'$ we find in the transition graph of F a subgraph as described in (c). This contradicts $L \in \mathcal{L}_{3/2}$. Hence F does not have pattern $\mathbb{L}_1^{\mathcal{L}}$, so $L \in \mathcal{C}_1^{\mathcal{L}}$.

Conversely, let $L \in \mathcal{C}_1^{\mathcal{L}}$ and let $F = (A, S, \delta, s_0, S')$ be the minimal dfa with $L(F) = L$. Assume that F has a subgraph in its transition graph as stated in (c). So let $u, v, w, z \in A^*$ with $\alpha(v) = \alpha(w)$ such that $\delta(s_0, u) = s_1 = \delta(s_1, v)$, $\delta(s_1, w) = s_2 = \delta(s_2, v)$, $\delta(s_1, z) \in S'$ and $\delta(s_2, z) \notin S'$ (cf. Figure 7). We want to obtain a contradiction to $L \in \mathcal{C}_1^{\mathcal{L}}$. Note with Definition 3.5 that it suffices to show that s_1 and s_2 are connected via some $p' \in \mathbb{L}_1^{\mathcal{L}}$. In order to achieve this, we want to provide the prerequisites of Lemma 4.3. Observe that $w \in A^+$ because otherwise $s_1 = s_2$ which is not possible. Since $\alpha(v) = \alpha(w)$ also $v \in A^+$. Let $l =_{\text{def}} v$ and $b =_{\text{def}} vw$. Then $p =_{\text{def}} (l, b) \in \mathbb{L}_0^{\mathcal{B}} = A^+ \times A^+$ and $\alpha(b) = \alpha(vw) = \alpha(v) = \alpha(l)$. Moreover, l and b start with the same letter and s_1, s_2 are connected via p. So we obtain from Lemma 4.3 that there is some $p' \in \mathbb{L}_1^{\mathcal{L}}$ such that s_1, s_2 are connected via p'. It follows that $L \notin \mathcal{C}_1^{\mathcal{L}}$, contradicting our assumption. Hence $L \in \mathcal{L}_{3/2}$ or $L = L' \cup \{\varepsilon\}$ for some $L' \in \mathcal{L}_{3/2}$ by (c). The latter is not possible since $L \subseteq A^+$ by definition of $\mathcal{C}_1^{\mathcal{L}}$.

Statement 5: The following is shown in [GS00].

(d) Let F be a dfa with $L(F) \subseteq A^+$. Then $L(F) \in \mathcal{B}_{3/2}$ if and only if F does not have a subgraph in its transition graph as depicted in Figure 10 with $m \geq 0$, $\tilde{u}_i, \tilde{w}_i \in A^+$, $\tilde{u}, \tilde{z}, \tilde{v}_i \in A^*$, and s^+ (s^-) being an accepting (rejecting) state.

This is [GS00, Theorem 6] (recall also by Theorem 2.20 that we talk about the same class of languages $\mathcal{B}_{3/2}$). We first equivalently restate the appearance of some $\mathbb{L}_1^{\mathcal{B}}$ as certain reachability conditions in the transition graph of a dfa. We show afterwards that this is equivalent to the appearance of the subgraph mentioned in (d). Let $F = (A, S, \delta, s_0, S')$ be a dfa and let $p \in \mathbb{L}_1^{\mathcal{B}}$. Then $s_1, s_2 \in S$ are connected via p if and only if

($*$) there are $m \geq 0$, $w_i, l_i, b_i \in A^+$ and $\tilde{r}_i, \tilde{s}_i, \tilde{q}_i, \tilde{r}'_i, \tilde{s}'_i \in S$ for $0 \leq i \leq m$ such that

1. $\tilde{r}_0 = \delta(s_1, w_0)$ and $\tilde{r}'_0 = \delta(s_2, w_0)$,
2. $\tilde{s}_m = s_1$ and $\tilde{s}'_m = s_2$,
3. $\delta(\tilde{s}_j, w_{j+1}) = \tilde{r}_{j+1}$ and $\delta(\tilde{s}'_j, w_{j+1}) = \tilde{r}'_{j+1}$ for $0 \leq j < m$,
4. $\delta(\tilde{r}_i, b_i) = \tilde{s}_i$ and $\delta(\tilde{r}'_i, b_i) = \tilde{s}'_i$ for $0 \leq i \leq m$,
5. $\tilde{q}_0 = \delta(s_1, w_0)$, $\tilde{q}_m = s_2$ and $\delta(\tilde{q}_j, w_{j+1}) = \tilde{q}_{j+1}$ for $0 \leq j < m$,
6. $\delta(s, l_i) = s$ for $1 \leq i \leq m$ where s is either $\tilde{r}_i, \tilde{s}_i, \tilde{q}_i, \tilde{r}'_i$ or $\tilde{s}'_i$.

The conditions ($*$) are displayed in Figure 8. Only a few observations are needed since this is a straightforward translation of Definition 3.4. To see the only-if–part let $m \geq 0$ and $p = (w_0, p_0, \ldots, w_m, p_m) \in \mathbb{L}_1^{\mathcal{B}}$ with $p_i = (l_i, b_i) \in \mathbb{L}_0^{\mathcal{B}} = A^+ \times A^+$ and $w_i \in A^+$ for $0 \leq i \leq m$. All statements can be verified with help of the witnessing states from Definition 3.4. For the if–part let $p =_{\text{def}} (w_0, p_0, \ldots, w_m, p_m)$ with $p_i =_{\text{def}} (l_i, b_i)$ for $0 \leq i \leq m$. Then $p_i \in \mathbb{L}_0^{\mathcal{B}}$ and $p \in \mathbb{L}_1^{\mathcal{B}}$. Just consider Definition 3.4 again and observe from statements 3 and 6 that for $0 \leq i \leq m$ the states $\tilde{r}_i, \tilde{s}_i$ and $\tilde{r}'_i, \tilde{s}'_i$ are connected via p_i, and from statement 6 that p_i appears at $\tilde{q}_i$.

Now we want to see that the occurrences of the subgraphs in Figures 8 and 10 imply one another. This is an easy transformation of patterns, which we carry out next. First suppose $L \notin \mathcal{B}_{3/2}$ and let $F = (A, S, \delta, s_0, S')$ be a dfa with $L(F) = L$. Then we find in the transition graph of F a subgraph as stated in (d) with $m \geq 0$, $\tilde{u}_i, \tilde{w}_j \in A^+$ and $\tilde{u}, \tilde{z}, \tilde{v}_i \in A^*$ for $1 \leq j \leq m+1$, $0 \leq i \leq m+1$, and s^+ (s^-) being an accepting (rejecting) state. It suffices to ensure the reachability conditions ($*$) to see that s_1, s_2 are connected via some $p \in \mathbb{L}_1^{\mathcal{B}}$, and hence $L \notin \mathcal{C}_1^{\mathcal{B}}$. Therefore, define

1. $m' =_{\text{def}} m + 1$,
2. $w_0 =_{\text{def}} \tilde{u}_0$ and $w_i =_{\text{def}} \tilde{w}_i$ for $1 \leq i \leq m'$,
3. $l_i =_{\text{def}} \tilde{u}_i$ for $0 \leq i \leq m'$, and
4. $b_i =_{\text{def}} \tilde{v}_i \tilde{u}_i$ for $0 \leq i \leq m'$.

Then $w_i, l_i \in A^+$ and with the last definition we also ensure that $b_i \in A^+$. Moreover, let

1. $\tilde{r}_0 =_{\text{def}} s_1$, $\tilde{r}'_0 =_{\text{def}} s_2$, and $\tilde{q}_0 =_{\text{def}} s_1$,
2. $\tilde{s}_i =_{\text{def}} \delta(s_1, \tilde{v}_0 \tilde{w}_1 \cdots \tilde{w}_i \tilde{v}_i)$ and $\tilde{s}'_i =_{\text{def}} \delta(s_2, \tilde{v}_0 \tilde{w}_1 \cdots \tilde{w}_i \tilde{v}_i)$ for $0 \leq i \leq m'$,
3. $\tilde{r}_j =_{\text{def}} \delta(s_1, \tilde{v}_0 \tilde{w}_1 \cdots \tilde{w}_{j-1} \tilde{v}_{j-1} \tilde{w}_j)$ and $\tilde{r}'_j =_{\text{def}} \delta(s_2, \tilde{v}_0 \tilde{w}_1 \cdots \tilde{w}_{j-1} \tilde{v}_{j-1} \tilde{w}_j)$ for $1 \leq j \leq m'$,
4. $\tilde{q}_j =_{\text{def}} \delta(s_1, \tilde{w}_1 \cdots \tilde{w}_j)$ for $1 \leq j \leq m'$.

Since $\delta(s_1, w_0) = \delta(s_1, u_0) = s_1 = \tilde{r}_0$, $\delta(s_2, w_0) = \delta(s_2, u_0) = s_2 = \tilde{r}'_0$ and $\delta(s_1, w_0) = \delta(s_1, u_0) = s_1 = \tilde{q}_0$ we can verify all conditions $(*)$. Note that $\delta(\tilde{r}_i, b_i) = \delta(\tilde{r}_i, \tilde{v}_i\tilde{u}_i) = \delta(\tilde{r}_i, \tilde{v}_i) = \tilde{s}_i$ and also $\delta(\tilde{r}'_i, b_i) = \delta(\tilde{r}'_i, \tilde{v}_i\tilde{u}_i) = \delta(\tilde{r}'_i, \tilde{v}_i) = \tilde{s}'_i$ for $0 \leq i \leq m'$.

Conversely, suppose $L \notin \mathcal{C}_1^{\mathcal{B}}$ and let $F = (A, S, \delta, s_0, S')$ be a dfa with $L(F) = L$. Then we find in the transition graph of F a subgraph as given in Figure 8, while we keep the notations from $(*)$. We want to show that we find a subgraph as given in Figure 10. Therefore, define

1. $m' =_{\mathrm{def}} m$,
2. $\tilde{u}_0 =_{\mathrm{def}} l_m$ and $\tilde{v}_0 =_{\mathrm{def}} l_m$,
3. $\tilde{w}_j =_{\mathrm{def}} w_{j-1}$, $\tilde{u}_j =_{\mathrm{def}} l_{j-1}$ and $\tilde{v}_j =_{\mathrm{def}} b_{j-1}$ for $1 \leq j \leq m' + 1$.

Then $\tilde{u}_i, \tilde{w}_i \in A^+$. Since $\delta(s_1, \tilde{u}_0) = \delta(s_1, l_m) = s_1$, $\delta(s_1, \tilde{v}_0) = \delta(s_1, l_m) = s_1$ and $\delta(s_2, \tilde{u}_0) = \delta(s_2, l_m) = s_2$, $\delta(s_2, \tilde{v}_0) = \delta(s_2, l_m) = s_2$ we find a subgraph as given in Figure 10 with these definitions. So by (d) we obtain that $L \notin \mathcal{B}_{3/2}$. This completes the proof of statement 5.

Statement 3, 6: This follows immediately from statement 1 and 4, together with Lemma 2.24, Theorem 3.15 and the monotony of the operators co and Pol. ❑

4.3 Pattern Classes are Starfree

Our pattern hierarchies exhaust the starfree languages.

Theorem 4.7. *It holds that* $\bigcup_{n\geq 0} \mathcal{C}_n^{\mathcal{B}} = \bigcup_{n\geq 0} \mathcal{C}_n^{\mathcal{L}} = SF$.

Proof. From the first two statements of Theorem 4.5 we get $\bigcup_{n\geq 0} \mathcal{C}_n^{\mathcal{B}} = \bigcup_{n\geq 0} \mathcal{C}_n^{\mathcal{L}}$. It holds that $\mathcal{C}_1^{\mathcal{L}} = \mathcal{L}_{3/2} \subseteq SF$ by Theorem 4.6 and Proposition 2.9. So we may apply Theorem 3.26 to obtain $\bigcup_{n\geq 0} \mathcal{C}_n^{\mathcal{L}} \subseteq SF$. Conversely, we get from Proposition 2.9 and Theorem 4.6 that $SF = \bigcup_{n\geq 1} \mathcal{B}_{n/2} \subseteq \bigcup_{n\geq 0} \mathcal{C}_n^{\mathcal{B}}$. ❑

4.4 The Hierarchy of Pattern Classes is Strict

We want to show the strictness of the two hierarchies $\{\mathcal{C}_n^{\mathcal{B}}\}$ and $\{\mathcal{C}_n^{\mathcal{L}}\}$ in a certain way, namely we take witnessing languages from [Tho84] that were used there to separate the classes of the dot–depth hierarchy. As remarked in [Tho84], these languages can also be used to show that the Straubing–Thérien hierarchy is strict. A first proof of strictness was given in [BK78] using similar languages. We could also do our separation here with these languages, but to facilitate the exposition we stick to [Tho84] since there the dot–depth hierarchy was defined exactly as we did here (namely, not taking ε into account).

We assume in this subsection that $A = \{a, b\}$. In fact, we will separate the instances of our hierarchies defined for this alphabet, see also Remark 4.14 below. Let us first define a family of patterns $\mathcal{W}_n$.

Definition 4.8. Let $F = (A, S, \delta, s_0, S')$ be a dfa and let $n \geq 1$. We say that F has pattern $\mathcal{W}_n$ if and only there are states $r_0, r_1, \ldots, r_n \in S$ such that for $0 \leq i \leq n-1$ it holds that $\delta(r_i, a) = r_{i+1}$ and $\delta(r_{i+1}, b) = r_i$.

Lemma 4.9. *Let $n \geq 2$. Then there exist $p, p' \in \mathbb{L}_{n-1}^{\mathcal{L}}$ such that for every dfa $F = (A, S, \delta, s_0, S')$ and for every occurrence $r_0, r_1, \ldots, r_n \in S$ of $\mathcal{W}_n$ in F it holds that*

1. *the states r_0 and r_1 are connected via p, and*
2. *the states r_1 and r_0 are connected via p'.*

Proof. We show the lemma by induction on n. For the induction base let $n = 2$ and define $\hat{p} =_{\mathrm{def}} (ab, a)$ and $\hat{p}' =_{\mathrm{def}} (ab, abb)$. Obviously, $\hat{p}, \hat{p}' \in \mathbb{L}_0^{\mathcal{B}}$. We can apply Lemma 4.3 to see that in fact there are $p, p' \in \mathbb{L}_1^{\mathcal{C}}$ such that if two states are connected via $\hat{p}$ or $\hat{p}'$ in a dfa F, then they are connected via p or p', respectively. Note that the definition of p and p' does not depend on F. So it suffices to do the following.

Let some states $r_0, r_1, r_2 \in S$ witness that some dfa $F = (A, S, \delta, s_0, S')$ has pattern $\mathcal{W}_1$. Then it holds that $\delta(r_0, ab) = \delta(r_1, b) = r_0$ and $\delta(r_1, ab) = \delta(r_2, b) = r_1$, so $\hat{p}$ and $\hat{p}'$ both appear at r_0 and r_1. Since $\delta(r_0, a) = r_1$ and $\delta(r_1, abb) = r_0$ we see that the states r_0 and r_1 are connected via $\hat{p}$, and that the states r_1 and r_0 are connected via $\hat{p}'$. This shows the induction base.

Now assume that there exist $\hat{p}, \hat{p}' \in \mathbb{L}_{n-1}^{\mathcal{C}}$ for some $n \geq 2$ having the properties stated in the lemma and we want to show the lemma for $n + 1$. Define $p =_{\mathrm{def}} (a, \hat{p}')$ and $p' =_{\mathrm{def}} (ab, \hat{p}, b, \lambda(\hat{p}))$. Obviously, $p, p' \in \mathbb{L}_n^{\mathcal{C}}$. Suppose we are given a dfa $F = (A, S, \delta, s_0, S')$ and states $r_0, r_1, \ldots, r_n, r_{n+1} \in S$ that witness an occurrence of pattern $\mathcal{W}_{n+1}$ in F. Since the states $r_0, \ldots, r_n$ and also the states $r_1, \ldots, r_{n+1}$ show that F has pattern $\mathcal{W}_n$ we can apply the induction hypothesis and obtain that

(a) r_0 and r_1 are connected via $\hat{p}$,

(b) r_1 and r_0 are connected via $\hat{p}'$,

(c) r_1 and r_2 are connected via $\hat{p}$ and

(d) r_2 and r_1 are connected via $\hat{p}'$.

Since $\hat{p}$ appears at r_0 and r_1 Lemma 3.8 provides us with the following.

(e) r_0 and r_0 are connected via $\lambda(\hat{p})$ and

(f) r_1 and r_1 are connected via $\lambda(\hat{p})$.

Let us verify that r_0 and r_1 are connected via p. Since $\delta(r_0, a) = r_1$ and because r_1 and r_0 are connected via $\hat{p}'$ by (b) we get that p appears at r_0. Similarly, since $\delta(r_1, a) = r_2$ and because r_2 and r_1 are connected via $\hat{p}'$ by (d) we get that p appears at r_1. Moreover, $\delta(r_0, a) = r_1$ and because $\hat{p}'$ appears at r_1 by (b) we obtain that r_0 and r_1 are connected via p.

Now we want so see that r_1 and r_0 are connected via p'. We get that p' appears at r_1 because $\delta(r_1, ab) = r_1$, the states r_1 and r_2 are connected via $\hat{p}$ by (c), $\delta(r_2, b) = r_1$, and since r_1 and r_1 are connected via $\lambda(\hat{p})$ by (f). Similarly, we get that p' appears at r_0 because $\delta(r_0, ab) = r_0$, the states r_0 and r_1 are connected via $\hat{p}$ by (a), $\delta(r_1, b) = r_0$, and since r_0 and r_0 are connected via $\lambda(\hat{p})$ by (e). Finally, because $\delta(r_1, ab) = r_1$ and $\hat{p}$ appears at r_1 by (a), and $\delta(r_1, b) = r_0$ and $\lambda(\hat{p})$ appears at r_0 by (e), we see that r_1 and r_0 are connected via p'. ❑

Next we recall the definition of a particular family of languages of A^+ from [Tho84]. Denote for $w \in A^+$ by $|w|_a$ ($|w|_b$) the number of occurrences of a (b, resp.) in w. Now define for $n \geq 1$ the language L_n to be the set of words $w \in A^+$ such that $|w|_a - |w|_b = n$ and for every prefix v of w it holds that $0 \leq (|v|_a - |v|_b) \leq n$. It was shown in [Tho84] that $L_n \in \mathcal{B}_n$ (these languages were denoted as L_n^+ there).

Definition 4.10. Let $n \geq 1$. Define $F_n =_{\mathrm{def}} (A, S, \delta, r_0, \{r_n\})$ with $S =_{\mathrm{def}} \{r_0, r_1, \ldots r_n, r^-\}$ and

- $\delta(r_i, a) = r_{i+1}$ and $\delta(r_{i+1}, b) = r_i$ for $0 \leq i \leq n - 1$,
- $\delta(r_n, a) = r^-$ and $\delta(r_0, b) = r^-$, and
- $\delta(r^-, a) = \delta(r^-, b) = r^-$.

It is easy to see that $L_n = L(F_n)$. We turn to our hierarchy classes.

Lemma 4.11. *Let $n \geq 1$. Then $L_n \in \mathcal{C}_n^{\mathcal{B}}$*

Proof. As already mentioned, we get from [Tho84] that $L_n \in \mathcal{B}_n$. So $L_n \in \mathcal{B}_{n+1/2}$ by Proposition 2.7 and by Theorem 4.6 we obtain $L_n \in \mathcal{C}_n^{\mathcal{B}}$. ❑

Lemma 4.12. *Let $n \geq 1$. Then $L_n \notin \mathcal{C}_n^{\mathcal{L}}$*

Proof. Let $n = 1$. Then we have to show that F_1 has pattern $\mathbb{L}_1^{\mathcal{L}}$. Let us first show that there is some $p \in \mathbb{L}_0^{\mathcal{B}}$ such that r_0 and r^- are connected via p. If we define $p = (ab, abb)$ we get $p \in \mathbb{L}_0^{\mathcal{B}}$ and since $\delta(r_0, ab) = r_0$, $\delta(r^-, ab) = r^-$ and $\delta(r_0, abb) = r^-$ we obtain that r_0 and r^- are connected via p. Now we apply Lemma 4.3 to see that in fact there is some $\hat{p} \in \mathbb{L}_1^{\mathcal{L}}$ such that the states r_0 and r^- are connected via $\hat{p}$. Finally, define $v =_{\text{def}} \varepsilon$ and $z =_{\text{def}} a$ to witness that F_1 has pattern $\mathbb{L}_1^{\mathcal{L}}$.

Now suppose $n \geq 2$. Then we have to show that F_n has pattern $\mathbb{L}_n^{\mathcal{L}}$. Note that F_n has pattern $\mathcal{W}_n$. So from Lemma 4.9 we obtain that there exists $\hat{p} \in \mathbb{L}_{n-1}^{\mathcal{L}}$ such that the states r_0 and r_1 are connected via $\hat{p}$. Now define $p =_{\text{def}} (ab, \hat{p}, b, \lambda(\hat{p}))$. Then $p \in \mathbb{L}_n^{\mathcal{L}}$ and we show as before that r_0 and r^- are connected via p. Note that $\hat{p}$ appears at r_0 and so we have by Lemma 3.8 that the states r_0 and r_0 are connected via $\lambda(\hat{p})$.

We get that p appears at r_0 because $\delta(r_0, ab) = r_0$, the states r_0 and r_1 are connected via $\hat{p}$, $\delta(r_1, b) = r_0$, and since r_0 and r_0 are connected via $\lambda(\hat{p})$. Clearly, by Proposition 3.12 we have that p appears at r^- because $\delta(r^-, a) = \delta(r^-, b) = r^-$. Finally, because $\delta(r_0, ab) = r_0$ and $\hat{p}$ appears at r_0, and $\delta(r_0, b) = r^-$ and $\lambda(\hat{p})$ appears at r^- also by Proposition 3.12, we see that r_1 and r^- are connected via p. Finally, define $v =_{\text{def}} \varepsilon$ and $z =_{\text{def}} a^n$ to witness that F_n has pattern $\mathbb{L}_n^{\mathcal{L}}$. ❑

So we obtain the following theorem.

Theorem 4.13. *Let $n \geq 1$. Then it holds that*

1. $\mathcal{C}_{n-1}^{\mathcal{B}} \subsetneq \mathcal{C}_n^{\mathcal{B}}$ *and*
2. $\mathcal{C}_{n-1}^{\mathcal{L}} \subsetneq \mathcal{C}_n^{\mathcal{L}}$.

Proof. Taking together Lemma 4.11 and Lemma 4.12 we see that $L_n \in \mathcal{C}_n^{\mathcal{B}} \backslash \mathcal{C}_n^{\mathcal{L}}$. By Theorem 4.5 we have $\mathcal{C}_{n-1}^{\mathcal{B}} \subseteq \mathcal{C}_n^{\mathcal{L}} \subseteq \mathcal{C}_n^{\mathcal{B}}$. It follows that L_n is a witness for $\mathcal{C}_{n-1}^{\mathcal{B}} \subsetneq \mathcal{C}_n^{\mathcal{B}}$, which is statement 1.

On the other hand we have by Theorem 4.5 that $\mathcal{C}_n^{\mathcal{L}} \subseteq \mathcal{C}_n^{\mathcal{B}} \subseteq \mathcal{C}_{n+1}^{\mathcal{L}}$. So L_n is a witness for $\mathcal{C}_n^{\mathcal{L}} \subsetneq \mathcal{C}_{n+1}^{\mathcal{L}}$. Note from Theorem 4.6 that $\mathcal{C}_0^{\mathcal{L}} = \mathcal{L}_{1/2}$ and $\mathcal{C}_1^{\mathcal{L}} = \mathcal{L}_{3/2}$ and that $\mathcal{L}_{1/2} \subsetneq \mathcal{L}_{3/2}$ is known, e.g., [SW98]. So we get also statement 2. ❑

Remark 4.14. *Suppose we deal with some alphabet A such that $|A| > 2$, e.g., $A = \{a, b, c_1, \cdots, c_n\}$ for some $n \geq 1$. If we define F_n such that $\delta(s, c_i) = r^-$ for $1 \leq i \leq n$ and for all $s \in S$, we still find the desired patterns and can show Lemma 4.12. This means on the language side that we intersect the expressions for L_n with $\{a, b\}^+ = A^+ \backslash \bigcup_{1 \leq i \leq n} A^* c_i A^* \in \text{co}\mathcal{B}_{1/2}$. A look at [Tho84] makes clear that this does not increase the dot–depth, so Lemma 4.11 also holds.*

4.5 Decidability of the Pattern Classes $\mathcal{C}_n^{\mathcal{B}}$ and $\mathcal{C}_n^{\mathcal{L}}$

Next we see that our hierarchies of pattern classes structure the class of starfree languages in a decidable way. We can determine the membership to a hierarchy class even in an efficient way.

Theorem 4.15. *Fix some $n \geq 0$. On input of a dfa F it is decidable in nondeterministic logarithmic space whether or not $L(F)$ is in $\mathcal{C}_n^{\mathcal{B}}$ ($\mathcal{C}_n^{\mathcal{L}}$, resp.).*

Proof. It holds that $\text{PATTERN}_{0,k}^{\mathcal{L}}, \text{PATTERN}_{0,k}^{\mathcal{B}} \in \text{NL}$ for each $k \geq 1$. To see this observe that due to the definition of the initial pattern the problems $\text{PATTERN}_{0,k}^{\mathcal{L}}$ and $\text{PATTERN}_{0,k}^{\mathcal{B}}$ are just reachability problems very similar to REACH_k, which can be solved in NL (cf. Lemma 3.30). Now the theorem follows from Corollary 3.33. ❑

Since membership to SF is decidable, this yields an algorithm to determine the minimal n such that $L(F)$ is in $\mathcal{C}_n^{\mathcal{B}}$ ($\mathcal{C}_n^{\mathcal{L}}$, resp.).

4.6 Lower Bounds and a Conjecture for Concatenation Hierarchies

We subsume the inclusion structure of concatenation and pattern hierarchies in Figure 3. Inclusions hold from bottom to top, and doubled lines stand for equality. Observe the structural similarities w.r.t. inclusion in each hierarchy.

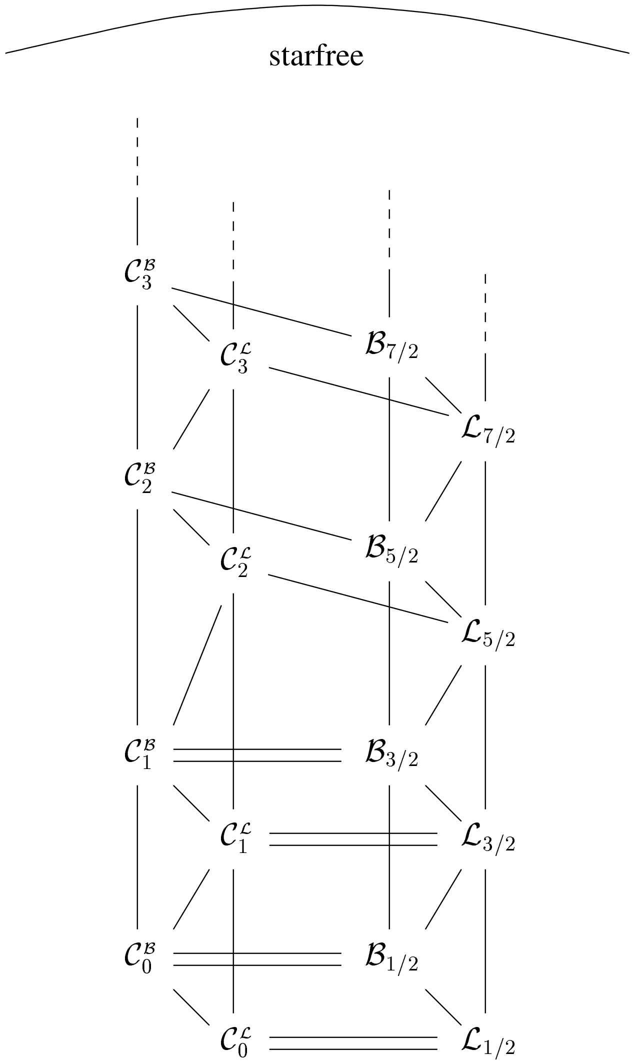


Figure 3: Concatenation Hierarchies and Forbidden Pattern Classes

In fact, the inclusions $\mathcal{B}_{n+1/2} \subseteq \mathcal{C}_n^{\mathcal{B}}$ and $\mathcal{L}_{n+1/2} \subseteq \mathcal{C}_n^{\mathcal{L}}$ establish a lower bound algorithm for the dot–depth of a given language. This follows from the fact that the pattern hierarchies are decidable (cf. Theorem 4.15). In the previous subsections, some necessary conditions as strictness and the upper bound SF were shown to hold for the pattern classes. So at this point we have no evidence against the following conjecture.

Conjecture 4.16. *For all $n \geq 0$ it holds that*

1. $\mathcal{B}_{n+1/2} = \mathcal{C}_n^{\mathcal{B}}$ *and*

2. $\mathcal{L}_{n+1/2} = \mathcal{C}_n^{\mathcal{L}}$.

Note that this conjectures an effective characterization of all levels $n + 1/2$ of the dot–depth hierarchy and the Straubing–Thérien hierarchy. As stated in Theorem 4.6 the conjecture is true for $n \in \{0, 1\}$. To get a better impression of this conjecture, the reader may compare Definitions 3.1 and 3.4 with Figures 4, 5, 6 and Figures 7, 8, 9.

If we look at the witnessing language L_n from Lemma 4.11 again we see that it is in $\mathcal{B}_{n+1/2}$ but not in $\mathcal{C}_n^{\mathcal{L}}$. So $\mathcal{B}_{n+1/2} \nsubseteq \mathcal{C}_n^{\mathcal{L}}$ which shows that the pattern class $\mathcal{C}_n^{\mathcal{L}}$ captures $\mathcal{L}_{n+1/2}$ but not $\mathcal{B}_{n+1/2}$. We may conclude that our pattern classes are not 'too big'.

5 $\mathcal{L}_{5/2}$ is decidable for two–letter alphabets

We show in this section that $\mathcal{C}_2^{\mathcal{L}} = \mathcal{L}_{5/2}$ in case of a two–letter alphabet. To do so, we prove a more general result: Whenever $\mathcal{B}_{n+1/2} = \mathcal{C}_n^{\mathcal{B}}$ for some $n \geq 1$ and arbitrary alphabets is known, then it holds that $\mathcal{L}_{n+3/2} = \mathcal{C}_{n+1}^{\mathcal{L}}$ in case of a two–letter alphabet. Due to the result $\mathcal{B}_{3/2} = \mathcal{C}_1^{\mathcal{B}}$ from [GS99], which is valid for arbitrary alphabets and restated in Theorem 4.6, it then follows in particular that $\mathcal{L}_{5/2} = \mathcal{C}_2^{\mathcal{L}}$ in the two–letter case. Let $A =_{\text{def}} \{a, b\}$ in this section.

We use the observation made in Proposition 2.3 to encode the alternating blocks of a's and b's in a word from A^+ over some larger alphabet. For this end, we look for the remainder of this section at some fixed but arbitrary minimal permutationfree dfa $F = (A, S, \delta, s_0, S')$ and set $r =_{\text{def}} |F|$. Nothing new happens in F if it reads more that r consecutive letters a's or b's from the input.

5.1 Changing the Alphabet

Let $A_F^a =_{\text{def}} \{a_1, \ldots, a_r\}$ and $A_F^b =_{\text{def}} \{b_1, \ldots, b_r\}$. We define an operation which maps words $w \in A^+$ to words $w_F \in (A_F)^+$ where $A_F =_{\text{def}} A_F^a \cup A_F^b$. Note that we can write every $w \in A^+$ as $w = w_1 w_2 \cdots w_n$ for some $n \geq 1$ and factors w_i of maximal length such that $w_i \in \{a\}^+ \cup \{b\}^+$. Call this the A-factorization of w and observe that this factorization is unique due to the maximality condition.

Definition 5.1. Let $w \in A^+$ and let $w = w_1 w_2 \cdots w_n$ for some $n \geq 1$ be the A-factorization of w. Then $w_F =_{\text{def}} c_1 c_2 \cdots c_n \in (A_F)^+$ with

$$c_i =_{\text{def}} \begin{cases} a_{\min\{l,r\}} & : \quad w_i = a^l \\ b_{\min\{l,r\}} & : \quad w_i = b^l \end{cases}$$

for $1 \leq i \leq n$ and $l \geq 1$.

We say that $c \in A_F$ has type a if $c \in A_F^a$ and it has type b if $c \in A_F^b$. Clearly, $w_i \in \{a\}^+$ if and only if c_i has type a and $w_i \in \{b\}^+$ if and only if c_i has type b. For later use, denote by $f(\mu)$ ($l(\mu)$) the type of the first (last, resp.) letter of $\mu \in (A_F)^+$.

Without further reference, we will make use of the following arguments. If $w = w_1 w_2 \cdots w_n$ for some $n \geq 1$ is the A-factorization of w then $w_F = (w_1)_F (w_2)_F \cdots (w_n)_F$, which follows directly from the above definition. Just observe that $w_F = c_1 c_2 \cdots c_n$ if and only if $(w_i)_F = c_i$ for $1 \leq i \leq n$. On the other hand, if we know for some $w \in A^+$ that $w_F = c_1 c_2 \cdots c_n$ for some $n \geq 1$ then the A-factorization of w has exactly n factors.

In the remainder of this subsection we give a number of easy to see Propositions that will be helpful in later proofs and that state basic properties. First observe that we encode the length $\leq r$ of w_i in c_i.

Proposition 5.2. *Let $w \in A^+$ and let $w = w_1 w_2 \cdots w_n$ be the A-factorization of w for some $n \geq 1$. Moreover, let $w_F = c_1 c_2 \cdots c_n$ for $c_i \in A_F$. If $c_i = a_j$ then $(w_i)_F = (a^j)_F$ and if $c_i = b_j$ then $(w_i)_F = (b^j)_F$ for $1 \leq i \leq n$ and $1 \leq j \leq r$.*

Proof. We show that $(w_i)_F = (a^j)_F$ if $c_i = a_j$, the other case is completely analogously. Because c_i has type a it must be that $w_i = a^l$ for some $l \geq 1$, and it holds that $j = \min\{l, r\}$ by definition. On one hand, we get $(a^j)_F = a_j$, on the other hand we have $(w_i)_F = (a^l)_F = a_{\min\{l,r\}} = a_j$. ❑

Next we see that this coding does what we intended, namely that F can not distinguish words having the same representation over A_F.

Proposition 5.3. *Let $w, v \in A^+$. If $w_F = v_F$ then $\delta(s, w) = \delta(s, v)$ for all $s \in S$.*

Proof. By assumption, $w_F = c_1c_2\cdots c_n = v_F$ for some $n \geq 1$ and $c_i \in A_F$. Let $w = w_1w_2\cdots w_n$ and let $v = v_1v_2\cdots v_n$ be the A-factorizations of w and v, respectively. Fix some $s \in S$ and some i with $1 \leq i \leq n$. We argue that $\delta(s, w_i) = \delta(s, v_i)$. Assume w.l.o.g. that c_i is of type a and hence $w_i, v_i \in \{a\}^+$. If $c_i = a_j$ with $1 \leq j \leq r-1$ then $w_i = a^j = v_i$ and hence $\delta(s, w_i) = \delta(s, v_i)$. If $c_i = a_r$ then $w_i = a^l$ and $v_i = a^{l'}$ with $l, l' \geq r$. By Proposition 2.3 we have that $\delta(s, a^l) = \delta(s, a^r) = \delta(s, a^{l'})$.

It follows inductively that $\delta(s, w_1w_2\cdots w_i) = \delta(s, v_1v_2\cdots v_i)$ for all $s \in S$ and $1 \leq i \leq n$. Hence, $\delta(s, w) = \delta(s, v)$. ❑

Corollary 5.4. *Let $w \in A^+$. Then it holds that*

$$w \in L(F) \iff v \in L(F)$$

for all $v \in A^+$ with $w_F = v_F$.

Proof. We only need to observe for $v \in A^+$ that from $w_F = v_F$ we have by Proposition 5.3 that $\delta(s_0, w) = \delta(s_0, v)$. ❑

Infinitely many words from $(A_F)^+$ can appear in the range of the mapping $w \mapsto w_F$. The maximality condition in A-factorizations leads to alternations of letter types in w_F.

Definition 5.5. Let $WF =_{\text{def}} \{ \mu \in (A_F)^+ \mid \mu = w_F \text{ for some } w \in A^+ \}$ be the set of well–formed words of $(A_F)^+$. For $\mu \in WF$ define $L_\mu =_{\text{def}} \{ v \in A^+ \mid v_F = \mu \}$.

By definition, none of the sets L_μ is empty.

Proposition 5.6. *Let $\mu = c_1c_2\cdots c_n \in (A_F)^+$ for some $n \geq 1$. Then $\mu \in WF$ if and only if $c_i \in A_F^a \Leftrightarrow c_{i+1} \in A_F^b$ for $1 \leq i < n$.*

Proof. As mentioned before, if $\mu = w_F$ for some $w \in A^+$ then the letters c_i alternate between A_F^a and A_F^b due to the maximality condition in the A-factorization of w. Conversely, we may take $w =_{\text{def}} w_1w_2\cdots w_n$ with $w_i =_{\text{def}} a^j$ if $c_i = a_j \in A_F^a$ and $w_i =_{\text{def}} b^j$ if $c_i = b_j \in A_F^b$. Then $w = w_1w_2\cdots w_n$ is an A-factorization of w with $(w_i)_F = c_i$. So $w_F = (w_1)_F(w_2)_F\cdots(w_n)_F = c_1c_2\cdots c_n$ and hence $w_F = \mu \in WF$. ❑

Clearly, factors of well–formed words are again well–formed.

Proposition 5.7. *Let $\mu = \mu_1\mu_2\mu_3 \in WF$ for some $\mu_1, \mu_3 \in (A_F)^*$ and $\mu_2 \in (A_F)^+$. Then $\mu_2 \in WF$.*

Proof. Let $\mu = c_1c_2\cdots c_n \in (A_F)^+$ for some $n \geq 1$ and suppose $\mu_2 = c_jc_{j+1}\cdots c_{j'}$ for some j, j' with $1 \leq j \leq j' \leq n$. Since $\mu \in WF$ we have by Proposition 5.6 that $c_i \in A_F^a \Leftrightarrow c_{i+1} \in A_F^b$ for $1 \leq i < n$. This holds in particular for $j \leq i < j'$. ❑

On the other hand, we can concatenate words from A^+ if the concatenation of their images is well–formed.

Proposition 5.8. *Let $w, v \in A^+$ such that $w_Fv_F \in WF$. Then $w_Fv_F = (wv)_F$.*

Proof. Suppose $w = w_1w_2\cdots w_n$ and $v = w_{n+1}w_{n+2}\cdots w_{n+m}$ for $n, m \geq 1$ are the A-factorizations of w and v, respectively. Then $w_F = c_1c_2\cdots c_n$ and $v_F = c_{n+1}c_{n+2}\cdots c_{n+m}$ for some $c_i \in A_F$ with $1 \leq i \leq n+m$. Moreover, it holds that $(w_i)_F = c_i$ for $1 \leq i \leq n+m$. Since $w_Fv_F \in WF$ we have by Proposition 5.6 that in particular c_n and c_{n+1} have different type. It follows that $w_n \in \{a\}^+ \Leftrightarrow w_{n+1} \in \{b\}^+$. So the A-factorization of wv is $wv = w_1\cdots w_nw_{n+1}\cdots w_{n+m}$. Hence, $(wv)_F = (w_1)_F\cdots(w_n)_F(w_{n+1})_F\cdots(w_{n+m})_F = c_1\cdots c_nc_{n+1}\cdots c_{n+m} = w_Fv_F$. ❑

Finally, let us denote for further reference that we can take factors of w that respect the A-factorization and obtain the respective factor of w_F.

Proposition 5.9. *Let $w \in A^+$ and let $w = w_1 w_2 \cdots w_n$ be the A-factorization of w for some $n \geq 1$. If $w_F = c_1 c_2 \cdots c_n$ for $c_i \in A_F$ then for all j, j' with $1 \leq j \leq j' \leq n$ it holds that $(w_j w_{j+1} \cdots w_{j'})_F = c_j c_{j+1} \cdots c_{j'}$.*

Proof. We have $(w_i)_F = c_i$ for $1 \leq i \leq n$ and so $c_j c_{j+1} \cdots c_{j'} = (w_j)_F (w_{j+1})_F \cdots (w_{j'})_F$. Since $w_j w_{j+1} \cdots w_{j'}$ is an A-factorization we have $(w_j)_F (w_{j+1})_F \cdots (w_{j'})_F = (w_j w_{j+1} \cdots w_{j'})_F$. ❑

Remark 5.10. *The mapping $w \mapsto w_F$ induces in a natural way an equivalence relation on A^+ that depends on F, i.e., define for $w, v \in A^+$ that $w \bowtie_F v$ if and only if $w_F = v_F$. Denote by $[w]_F$ the equivalence class including w and by $A^+/_{\bowtie_F}$ the set of all equivalence classes. Then $A^+/_{\bowtie_F}$ and WF are isomorphic and $\bowtie_F$ has an infinite index on A^+. Moreover, it holds that $[w]_F = L_{w_F}$ for all $w \in A^+$. In fact, Proposition 5.3 shows that F can not distinguish equivalent words, so $\bowtie_F$ is a refinement of the syntactical congruence.*

5.2 Automata Construction

We can take words from $(A_F)^+$ as inputs to a dfa $\widetilde{F}$ which respects the acceptance behaviour of F. Moreover, we want the automaton to reject, whenever it finds two consecutive letters of the same type in the input (see Proposition 5.6). Therefore, we make the following canonical construction. We define

$$\widetilde{F} =_{\text{def}} (A_F, \widetilde{S}, \widetilde{\delta}, (s_0, \varepsilon), \widetilde{S}')$$

with

a. $\widetilde{S} =_{\text{def}} \{S \times \{a, b\}\} \cup \{(s_0, \varepsilon), \bot\}$,

b. $\widetilde{S}' =_{\text{def}} S' \times \{a, b\}$,

c. $\widetilde{\delta}(\bot, c) =_{\text{def}} \bot$ for all $c \in A_F$,

d. $\widetilde{\delta}((s_0, \varepsilon), a_j) =_{\text{def}} (\delta(s_0, a^j), a)$ for all $a_j \in A_F^a$,

e. $\widetilde{\delta}((s_0, \varepsilon), b_j) =_{\text{def}} (\delta(s_0, b^j), b)$ for all $b_j \in A_F^b$,

f. $\widetilde{\delta}((s, a), a_j) =_{\text{def}} \bot$ for all $a_j \in A_F^a$ and $s \in S$,

g. $\widetilde{\delta}((s, a), b_j) =_{\text{def}} (\delta(s, b^j), b)$ for all $b_j \in A_F^b$ and $s \in S$,

h. $\widetilde{\delta}((s, b), a_j) =_{\text{def}} (\delta(s, a^j), a)$ for all $a_j \in A_F^a$ and $s \in S$ and

i. $\widetilde{\delta}((s, b), b_j) =_{\text{def}} \bot$ for all $b_j \in A_F^b$ and $s \in S$.

We set $\widetilde{L} =_{\text{def}} L(\widetilde{F})$. Note that we simulate F in the first components and store the type of the last input letter in the second components of the states of $\widetilde{F}$. Let us make the relation between F and $\widetilde{F}$ precise.

Proposition 5.11. *Let $w \in A^+$ and $\delta(s, w) = s'$ for some $s, s' \in S$. Let $t_f \in A$ ($t_l \in A$) be the first (last, respectively) letter of w. For $t' \in A \cup \{\varepsilon\}$ with $t' \neq t_f$ and $(s, t') \in \widetilde{S}$ it holds that $\widetilde{\delta}((s, t'), w_F) = (s', t_l)$.*

Proof. Let $w = w_1 w_2 \cdots w_n$ with $n \geq 1$ be the A-factorization of w. Moreover, let $w_F = c_1 c_2 \cdots c_n$ and define t_i to be the type of c_i. Hence, $t_1 t_2 \cdots t_n$ consists of alternating a's and b's, and in particular, $t_1 = t_f$ and $t_n = t_l$. For $1 \leq i \leq n$ and $1 \leq j \leq r$ define $v_i =_{\text{def}} a^j$ if $c_i = a_j$ and $v_i =_{\text{def}} b^j$ if $c_i = b_j$. By Proposition 5.2 we have $(w_i)_F = (v_i)_F$ and so $\delta(p, w_i) = \delta(p, v_i)$ for all $p \in S$ by Proposition 5.3. Assume that $t' \in A \cup \{\varepsilon\}$ with $t' \neq t_f$ and $(s, t') \in \widetilde{S}$. We show inductively that $\widetilde{\delta}((s, t'), c_1 \cdots c_i) = (\delta(s, w_1 \cdots w_i), t_i)$ for $1 \leq i \leq n$.

Induction base. Let $i = 1$ and assume w.l.o.g. that $t_1 = t_f = a$. Hence, $c_1 = a_j$ and $v_1 = a^j$ for some $1 \leq j \leq r$, and $t' \in \{b, \varepsilon\}$ by assumption. First suppose $t' = \varepsilon$. Since we require that $(s, t') \in \widetilde{S}$ we only make an assertion in case $s = s_0$. We conclude from Proposition 2.3 and d. in the definition of $\widetilde{F}$ that

$$\begin{aligned}
\widetilde{\delta}((s, t'), c_1) &= \widetilde{\delta}((s_0, \varepsilon), a_j) \\
&= (\delta(s_0, a^j), a) \\
&= (\delta(s_0, v_1), a) \\
&= (\delta(s_0, w_1), t_1).
\end{aligned}$$

Now assume that $t' = b$. Then we see with Proposition 2.3 and h. in the definition of $\widetilde{F}$ that

$$\begin{aligned}
\widetilde{\delta}((s, t'), c_1) &= \widetilde{\delta}((s, b), a_j) \\
&= (\delta(s, a^j), a) \\
&= (\delta(s, v_1), a) \\
&= (\delta(s, w_1), t_1).
\end{aligned}$$

The case of $t_1 = t_f = b$ is completely analogously and involves e. and g. from the definition of $\widetilde{F}$.

Induction step. Suppose we have $\widetilde{\delta}((s, t'), c_1 \cdots c_i) = (\delta(s, w_1 \cdots w_i), t_i)$ for some i with $1 \leq i < n$ and we want to show this for $i + 1$. Let us assume w.l.o.g. that $t_i = a$. Hence, $c_i = a_j$, $v_i = a^j$ and $t_{i+1} = b$, $c_{i+1} = b_{j'}$, $v_{i+1} = b^{j'}$ for some $1 \leq j, j' \leq r$.

We conclude with the hypothesis, with Proposition 2.3 and with g. in the definition of $\widetilde{F}$ that for $1 \leq i < n$ we have

$$\begin{aligned}
\widetilde{\delta}((s, t'), c_1 \cdots c_i c_{i+1}) &= \widetilde{\delta}(\widetilde{\delta}((s, t'), c_1 \cdots c_i), c_{i+1}) \\
&= \widetilde{\delta}((\delta(s, w_1 \cdots w_i), t_i), c_{i+1}) \\
&= \widetilde{\delta}((\delta(s, w_1 \cdots w_i), a), b_{j'}) \\
&= (\delta(\delta(s_0, w_1 \cdots w_i), b^{j'}), b) \\
&= (\delta(\delta(s_0, w_1 \cdots w_i), v_{i+1}), b) \\
&= (\delta(\delta(s_0, w_1 \cdots w_i), w_{i+1}), t_{i+1}) \\
&= (\delta(s_0, w_1 \cdots w_i w_{i+1}), t_{i+1}).
\end{aligned}$$

The case of $t_i = b$ is completely analogously and involves h. from the definition of $\widetilde{F}$. This completes the induction and it follows that

$$\widetilde{\delta}((s, t'), w_F) = \widetilde{\delta}((s, t'), c_1 \cdots c_n) = (\delta(s, w_1 \cdots w_n), t_n) = (\delta(s, w), t_l) = (s', t_l).$$

❑

Proposition 5.12. *Suppose $\widetilde{\delta}((s, t), \mu) = (s', t')$ for $\mu \in (A_F)^+$ and $(s, t), (s', t') \in \widetilde{S}$ with $s, s' \in S$, $t \in A \cup \{\varepsilon\}$ and $t' \in A$. Then $\mu \in WF$ and $\delta(s, w) = s'$ for all $w \in L_\mu$.*

Proof. Let $\mu = c_1c_2\cdots c_n$ for some $n \geq 1$ and define again t_i to be the type of c_i. Note that if $t_1 = t$ then $\widetilde{\delta}((s,t),c_1) = \perp$ due to f. and i. in the definition of $\widetilde{F}$, and $\widetilde{\delta}((s,t),\mu) = \perp$ due to c. So $t_1 \neq t$.

Now assume that $\mu \notin WF$. Then we get from Proposition 5.6 that there is some minimal i with $1 \leq i < n$ such that $c_i, c_{i+1} \in A_F^a$ or $c_i, c_{i+1} \in A_F^b$. So $t_i = t_{i+1}$ and $c_1c_2\cdots c_i \in WF$. Hence, there is some $v \in A^+$ with first letter t_1 and last letter t_i such that $v_F = c_1\cdots c_i$ and there is some $s'' \in S$ with $\delta(s,v) = s''$. Because $t_1 \neq t$ we have all prerequisites of Proposition 5.11 and obtain from it that $\widetilde{\delta}((s,t),c_1c_2\cdots c_i) = (s'',t_i)$. From f. and i. in the definition of $\widetilde{F}$ we see that $\widetilde{\delta}((s'',t_i),c_{i+1}) = \perp$ since $t_i = t_{i+1}$, and from c. we get $\widetilde{\delta}((s,t),\mu) = \perp$, a contradiction. This shows that $\mu \in WF$.

Now let $w \in L_\mu$. Then $w_F = \mu$ and w has first letter t_1 and last letter t_n. There is some $s'' \in S$ such that $\delta(s,w) = s''$. Since we know that $t_1 \neq t$ we can apply Proposition 5.11 again and obtain

$$(s',t') = \widetilde{\delta}((s,t),\mu) = \widetilde{\delta}((s,t),w_F) = (s'',t_n).$$

So $\delta(s,w) = s'' = s'$. ❑

Since $\widetilde{S}' \subseteq \widetilde{S}\backslash\{\perp\}$ we get from Proposition 5.12 that $\widetilde{L} \subseteq WF$. More precisely, we have the following Proposition.

Proposition 5.13. *It holds that* $\widetilde{L} = \{\, w_F \mid w \in L \,\}$.

Proof. Suppose $\mu \in \widetilde{L}$. Then $\widetilde{\delta}((s_0,\varepsilon),\mu) = (s,d) \in \widetilde{S}'$ for some $s \in S'$ and $d \in A$ by b. in the definition of $\widetilde{F}$. We can apply Proposition 5.12 to see $\mu \in WF$. So there exists some $w \in A^+$ with $\mu = w_F$. Again by Proposition 5.12 we obtain $w \in L$ from $\delta(s_0,w) = s$.

Conversely, suppose $w \in L$. Then $\delta(s_0,w) = s$ for some $s \in S'$ and by Proposition 5.11 we get $\widetilde{\delta}((s_0,\varepsilon),w_F) = (s,t) \in \widetilde{S}'$ for some $t \in A$. So $w_F \in \widetilde{L}$. ❑

5.3 Transformation of Patterns

Let $n \geq 0$. We show that if $\widetilde{F}$ has pattern $\mathbb{L}_n^{\mathcal{B}}$ then F even has pattern $\mathbb{L}_{n+1}^{\mathcal{C}}$. The following lemma shows this for all patterns connecting some states not equal $\perp$. This ensures that the pattern contains only well–formed words. In the following, we consider $\mathbb{L}_n^{\mathcal{B}}$ with respect to A_F, and $\mathbb{L}_n^{\mathcal{C}}$ with respect to A.

Lemma 5.14. *Let $n \geq 0$ and $\tilde{p} \in \mathbb{L}_n^{\mathcal{B}}$ such that $(s,t),(s',t')$ are connected via $\tilde{p}$ in $\widetilde{F}$ for some $(s,t),(s',t') \in \widetilde{S}$ with $s,s' \in S$ and $t,t' \in A \cup \{\varepsilon\}$. Then there exists some $p \in \mathbb{L}_{n+1}^{\mathcal{C}}$ such that for all $(s_1,t_1),(s_2,t_2),(s_3,t_3) \in \widetilde{S}$ with $s_1,s_2,s_3 \in S$ and $t_1,t_2,t_3 \in A \cup \{\varepsilon\}$ the following holds.*

1. *If $\tilde{p}$ appears at (s_1,t_1) in $\widetilde{F}$, then p appears at s_1 in F.*
2. *If $(s_2,t_2),(s_3,t_3)$ are connected via $\tilde{p}$ in $\widetilde{F}$, then s_2,s_3 are connected via p in F.*

Proof. The proof is by induction on n.

Induction base. Let $n = 0$. Then $\tilde{p} = (\nu,\mu) \in \mathbb{L}_0^{\mathcal{B}}$ for some $\nu,\mu \in (A_F)^+$. Since $(s,t),(s',t') \in \widetilde{S}$ are connected via $\tilde{p}$ it follows that $\widetilde{\delta}((s,t),\nu^i) = (s,t)$ for all $i \geq 1$ and $\widetilde{\delta}((s,t),\mu) = (s',t')$. We obtain from Proposition 5.12 that $\nu^i,\mu \in WF$ for $i \geq 1$. So we can consider L_ν and L_μ, and there are $v \in L_\nu$ and $w \in L_\mu$ with $(v)_F = \nu$ and $(w)_F = \mu$. Note that $v,w \in A^+$ and define $p' =_{\mathrm{def}} (v,vw)$. Then $p' \in \mathbb{L}_0^{\mathcal{B}}$ (w.r.t. A), and v and vw start with the same letter. Moreover, it holds that $\alpha(v) = A$. To see this suppose $\alpha(v) \neq A$. Then all letters of ν have the same type $\alpha(v)$ and $\nu^2 \notin WF$, a contradiction. It follows that $\alpha(vw) \subseteq \alpha(v) = A$. Now Lemma 4.3 provides some $p \in \mathbb{L}_1^{\mathcal{C}}$ such that $\overline{p}^\circ = v^{|vw|}$ and if any two states in F are connected via p' then they are also connected via p.

To see statement 1 suppose $\tilde{p}$ appears at (s_1,t_1) in $\widetilde{F}$. Then $\widetilde{\delta}((s_1,t_1),\nu) = (s_1,t_1)$ and we get from Proposition 5.12 that $\delta(s_1,v) = s_1$. So $\delta(s_1,\overline{p}^\circ) = s_1$ and Proposition 4.2 shows that p appears at s_1 in F.

For statement 2 assume that $(s_2,t_2),(s_3,t_3)$ are connected via $\tilde{p}$ in $\widetilde{F}$. By definition, $\widetilde{\delta}((s_2,t_2),\nu) = (s_2,t_2)$, $\widetilde{\delta}((s_3,t_3),\nu) = (s_3,t_3)$ and $\widetilde{\delta}((s_2,t_2),\mu) = (s_3,t_3)$. We get from Proposition 5.12 that $\delta(s_2,v) = s_2$, $\delta(s_3,v) = s_3$ and $\delta(s_2,vw) = \delta(s_2,w) = s_3$. So s_2,s_3 are connected via p' in F and we have already seen that they are also connected via p.

Induction step. Suppose the lemma holds for some $n \geq 0$ and we want to show that is also holds for $n+1$. So let $\tilde{p} = (\mu_0,\tilde{p}_0,\ldots,\mu_m,\tilde{p}_m) \in \mathbb{L}^{\mathcal{B}}_{n+1}$ with $\mu_i \in (A_F)^+$ and $\tilde{p}_i \in \mathbb{L}^{\mathcal{B}}_n$ for $0 \leq i \leq m$. By assumption, there are $(s,t),(s',t') \in \widetilde{S}$ which are connected via $\tilde{p}$, and in particular, $\tilde{p}$ appears at (s,t). So there are states $\tilde{r}_i, \tilde{g}_i \in \widetilde{S}$ for $0 \leq i \leq m$ such that

- $\tilde{r}_0 = \widetilde{\delta}((s,t),\mu_0)$, $\tilde{g}_m = (s,t)$,
- $\widetilde{\delta}(\tilde{g}_j,\mu_{j+1}) = \tilde{r}_{j+1}$ for $0 \leq j < m$, and
- the states $\tilde{r}_i,\tilde{g}_i$ are connected via $\tilde{p}_i$ for $0 \leq i \leq m$.

Note that from Lemma 3.6 we obtain that for $\overline{\tilde{p}}^{\circ} = \mu_0\overline{\tilde{p}_0}\cdots\mu_m\overline{\tilde{p}_m}$ we have $\widetilde{\delta}((s,t),\overline{\tilde{p}}^{\circ}) = (s,t)$ and $\widetilde{\delta}(\tilde{r}_i,\overline{\tilde{p}_i}) = \tilde{g}_i$ for $0 \leq i \leq m$. Because of c. in the definition of $\widetilde{F}$ and since $(s,t) \neq \perp$ it follows that the states $\tilde{r}_i,\tilde{g}_i$ for $0 \leq i \leq m$ are not $\perp$. So we can rewrite these states as $\tilde{r}_i = (r_i,t_{r_i})$ and $\tilde{g}_i = (g_i,t_{g_i})$ for $0 \leq i \leq m$, and for suitable $r_i,g_i \in S$ and $t_{r_i},t_{g_i} \in A$. Due to the construction of $\widetilde{F}$ the latter are not ε. Now the induction hypothesis provides for each $\tilde{p}_i$ some $p_i \in \mathbb{L}^{\mathcal{L}}_{n+1}$ such that for all $(s'_1,t'_1),(s'_2,t'_2),(s'_3,t'_3) \in \widetilde{S}$ with $s'_1,s'_2,s'_3 \in S$ and $t'_1,t'_2,t'_3 \in A \cup \{\varepsilon\}$ the following holds.

(H1) If $\tilde{p}_i$ appears at (s'_1,t'_1) in $\widetilde{F}$, then p_i appears at s'_1 in F.

(H2) If $(s'_2,t'_2),(s'_3,t'_3)$ are connected via $\tilde{p}_i$ in $\widetilde{F}$, then s'_2,s'_3 are connected via p_i in F.

Moreover, we get from Proposition 5.12 that $\overline{\tilde{p}}^{\circ} \in \mathit{WF}$. So with Proposition 5.7 we see that $\mu_i \in \mathit{WF}$ for $0 \leq i \leq m$. In particular, there are $w_i \in L_{\mu_i}$ such that $(w_i)_F = \mu_i$ for $0 \leq i \leq m$. Define now $p =_{\text{def}} (w_0,p_0,\ldots,w_m,p_m)$ and observe that $p \in \mathbb{L}^{\mathcal{L}}_{n+2}$.

We turn to statement 1. Suppose $\tilde{p}$ appears at (s_1,t_1) in $\widetilde{F}$. Then there are states $\tilde{r}'_i$, $\tilde{g}'_i \in \widetilde{S}$ with $0 \leq i \leq m$ such that

- $\tilde{r}'_0 = \widetilde{\delta}((s_1,t_1),\mu_0)$, $\tilde{g}'_m = (s_1,t_1)$,
- $\widetilde{\delta}(\tilde{g}'_j,\mu_{j+1}) = \tilde{r}'_{j+1}$ for $0 \leq j < m$, and
- the states $\tilde{r}'_i,\tilde{g}'_i$ are connected via $\tilde{p}_i$ for $0 \leq i \leq m$.

With the same argument as above, we can rewrite these states as $\tilde{r}'_i = (r'_i,t_{r'_i})$ and $\tilde{g}'_i = (g'_i,t_{g'_i})$ for $0 \leq i \leq m$, and for suitable $r'_i,g'_i \in S$ and $t_{r'_i},t_{g'_i} \in A$. From Proposition 5.12 and by the induction hypothesis (H2) we obtain that

- $r'_0 = \delta(s_1,w_0)$, $g'_m = s_1$,
- $\delta(g'_j,w_{j+1}) = r'_{j+1}$ for $0 \leq j < m$, and
- the states r'_i,g'_i are connected via p_i for $0 \leq i \leq m$.

This shows that p appears at s_1 in F.

Next we want to prove statement 2. Therefore, assume that $(s_2,t_2),(s_3,t_3)$ are connected via $\tilde{p}$ in $\widetilde{F}$. Then $\tilde{p}$ appears at (s_2,t_2) and also at (s_3,t_3), and there are states $\tilde{q}_i \in \widetilde{S}$ with $0 \leq i \leq m$ such that

- $\tilde{q}_0 = \widetilde{\delta}((s_2,t_2),\mu_0)$,
- $\tilde{q}_m = (s_3,t_3)$,

- $\widetilde{\delta}(\tilde{q}_j, \mu_{j+1}) = \tilde{q}_{j+1}$ for $0 \le j < m$, and
- $\tilde{p}_i$ appears at state $\tilde{q}_i$ for $0 \le i \le m$.

As before, none of the states $\tilde{q}_i$ for $0 \le i \le m$ can be $\perp$, so we can rewrite them as $\tilde{q}_i = (q_i, t_{q_i})$ for $0 \le i \le m$, and for suitable $q_i \in S$ and $t_{q_i} \in A$. From Proposition 5.12 and by the induction hypothesis (H1) we obtain that

- $q_0 = \delta(s_2, w_0)$,
- $q_m = s_3$,
- $\delta(q_j, w_{j+1}) = q_{j+1}$ for $0 \le j < m$, and
- p_i appears at state q_i for $0 \le i \le m$.

Together with the appearance of p at s_2 and at s_3, which follows from statement 1, this shows that s_2, s_3 are connected via p in F. ❑

Lemma 5.15. *Let $n \ge 1$. Suppose $\widetilde{F}$ has pattern $\mathbb{L}_n^{\mathcal{B}}$ witnessed by $\tilde{p} \in \mathbb{L}_n^{\mathcal{B}}$ and $\eta, \zeta \in (A_F)^+$ such that*

- $\widetilde{\delta}((s_0, \varepsilon), \eta) = \tilde{s}_1$,
- $\widetilde{\delta}(\tilde{s}_1, \zeta) \in \widetilde{S}'$,
- $\widetilde{\delta}(\tilde{s}_2, \zeta) \notin \widetilde{S}'$, *and*
- *the states $\tilde{s}_1, \tilde{s}_2$ are connected via $\tilde{p}$.*

Then $\eta \cdot \zeta \in WF$ and $\eta \cdot \overline{\tilde{p}} \cdot \zeta \in WF$.

Proof. Let $\tilde{p} = (\mu_0, \tilde{p}_0, \ldots, \mu_m, \tilde{p}_m)$ for some $m \ge 0$ with $\mu_i \in (A_F)^+$ and $\tilde{p}_i \in \mathbb{L}_{n-1}^{\mathcal{B}}$. Recall that $\overline{\tilde{p}}^{\circ} = \mu_0 \overline{\tilde{p}_0} \cdots \mu_m \overline{\tilde{p}_m}$ and that $\widetilde{\delta}(\tilde{s}_1, \overline{\tilde{p}}^{\circ}) = \tilde{s}_1$ by Lemma 3.6. So $\eta \cdot \zeta, \eta \cdot \overline{\tilde{p}}^{\circ} \cdot \zeta \in \widetilde{L}$ and hence, both words are in WF by Proposition 5.13. Because $n \ge 1$ we can look at the occurrences of the $\tilde{p}_i$ in $\widetilde{F}$ for $0 \le i \le m$ witnessing that $\tilde{p}$ appears at $\tilde{s}_1$. Again by Lemma 3.6 we see that if $\tilde{p}_i$ appears at some state in $\widetilde{F}$ then this state has a $\overline{\tilde{p}_i}^{\circ}$-loop. So we have

$$\eta \cdot \mu_0 \cdot \overline{\tilde{p}_0} \cdot \mu_1 \cdot \overline{\tilde{p}_1} \cdots \mu_m \cdot \overline{\tilde{p}_m} \cdot \zeta \in \widetilde{L} \subseteq WF \text{ and} \tag{18}$$
$$\eta \cdot \mu_0 \cdot \overline{\tilde{p}_0}^{\circ} \cdot \overline{\tilde{p}_0} \cdot \overline{\tilde{p}_0}^{\circ} \cdot \mu_1 \cdot \overline{\tilde{p}_1}^{\circ} \cdot \overline{\tilde{p}_1} \cdot \overline{\tilde{p}_1}^{\circ} \cdots \mu_m \cdot \overline{\tilde{p}_m}^{\circ} \cdot \overline{\tilde{p}_m} \cdot \overline{\tilde{p}_m}^{\circ} \cdot \zeta \in \widetilde{L} \subseteq WF. \tag{19}$$

Note that none of these factors is the empty word (since $\mathcal{B} = A^+ \times A^+$), and that the types of every two consecutive letters alternate. To argue that $\eta \cdot \overline{\tilde{p}} \cdot \zeta = \eta \cdot \mu_0 \mu_1 \cdots \mu_m \cdot \zeta \in WF$ we need to observe that also in this word the types of every two consecutive letters alternate. It suffices to show for $0 \le i \le m$ that $l(\mu_i) = l(\overline{\tilde{p}_i})$ which can be seen from Eq. (18). We obtain from Eq. (19) that $l(\overline{\tilde{p}_i}) \ne f(\overline{\tilde{p}_i}^{\circ})$. Now assume $l(\mu_i) \ne l(\overline{\tilde{p}_i})$ for some i with $0 \le i \le m$. Then $l(\mu_i) = f(\overline{\tilde{p}_i}^{\circ})$, a contradiction to Eq. (19). ❑

Theorem 5.16. *Let $n \ge 1$. If $\widetilde{F}$ has pattern $\mathbb{L}_n^{\mathcal{B}}$ then F has pattern $\mathbb{L}_{n+1}^{\mathcal{C}}$.*

Proof. Suppose that $\widetilde{F}$ has pattern $\mathbb{L}_n^{\mathcal{B}}$. So there exist $\tilde{s}_1, \tilde{s}_2 \in \widetilde{S}, \eta, \zeta \in (A_F)^*, \tilde{p} \in \mathbb{L}_n^{\mathcal{B}}$ such that

- $\widetilde{\delta}((s_0, \varepsilon), \eta) = \tilde{s}_1$,
- $\widetilde{\delta}(\tilde{s}_1, \zeta) \in \widetilde{S}'$,
- $\widetilde{\delta}(\tilde{s}_2, \zeta) \notin \widetilde{S}'$, and

- the states $\tilde{s}_1, \tilde{s}_2$ are connected via $\tilde{p}$.

Set $\tilde{s}^+ =_{\text{def}} \widetilde{\delta}(\tilde{s}_1, \zeta)$ and $\tilde{s}^- =_{\text{def}} \widetilde{\delta}(\tilde{s}_2, \zeta)$. We have $\eta \neq \varepsilon$, otherwise there is a non–empty loop $\overline{\tilde{p}}^{\circ}$ at $(s_0, \varepsilon) = \tilde{s}_1$, which is not possible by construction. Moreover, we may suppose that $\zeta \neq \varepsilon$ since otherwise we take $\zeta' =_{\text{def}} \overline{\tilde{p}}^{\circ}$ and obtain with Lemma 3.6 that $\widetilde{\delta}(\tilde{s}_1, \zeta') = \tilde{s}_1$ and $\widetilde{\delta}(\tilde{s}_2, \zeta') = \tilde{s}_2$. From Lemma 5.15 we get that $\eta \cdot \zeta \in WF$ and $\eta \cdot \overline{\tilde{p}} \cdot \zeta \in WF$. In particular, Proposition 5.7 says that any prefix of these words is in WF. Let $u, z \in A^+$ such that $u_F = \eta$ and $z_F = \zeta$.

We want to argue that none of the (not necessarily strict) prefixes of $\eta \cdot \zeta$ and $\eta \cdot \overline{\tilde{p}} \cdot \zeta$ leads to $\perp$ in $\widetilde{F}$. To see this let $\mu \in (A_F)^+$ be an arbitrary such prefix and let $w \in A^+$ such that $w_F = \mu$. Clearly, if we give w as an input to F we end up in some state $s \in S$. Now Proposition 5.11 tells us that $\widetilde{\delta}((s_0, \varepsilon), w_F) = \widetilde{\delta}((s_0, \varepsilon), \mu) = (s, t) \neq \perp$ for some $t \in A$. So we can rewrite the occurring states as

- $\tilde{s}_1 = (s_1, t_{s_1})$ and $\tilde{s}_2 = (s_2, t_{s_2})$,
- $\tilde{s}^+ = (s^+, t_{s^+})$ and $\tilde{s}^- = (s^-, t_{s^-})$

for suitable $s_1, s_2, s^+, s^- \in S$ with $s^+ \in S'$, $s^- \notin S'$, and $t_{s_1}, t_{s_2}, t_{s^+}, t_{s^-} \in A$. From Proposition 5.12 and by Lemma 5.14 we obtain that

- $\delta(s_0, u) = s_1$,
- $\delta(s_1, z) = s^+ \in S'$,
- $\delta(s_2, z) = s^- \notin S'$, and
- the states s_1, s_2 are connected via some $p \in \mathbb{L}^{\mathcal{L}}_{n+1}$.

This shows that F has pattern $\mathbb{L}^{\mathcal{L}}_{n+1}$. ❑

Corollary 5.17. *Let $n \geq 1$. If F does not have pattern $\mathbb{L}^{\mathcal{L}}_{n+1}$ then $\widetilde{L} \in \mathcal{C}^{\mathcal{B}}_n$.*

Proof. By Theorem 5.16 we know that if F does not have pattern $\mathbb{L}^{\mathcal{L}}_{n+1}$ then $\widetilde{F}$ does not have pattern $\mathbb{L}^{\mathcal{B}}_n$. So $L(\widetilde{F}) = \widetilde{L} \in \mathcal{C}^{\mathcal{B}}_n$, since $\widetilde{L} \subseteq (A_F)^+$. ❑

5.4 Transformation of Expressions

Recall that we still deal with the two–letter alphabet A and our fixed permutationfree dfa F. We want to show in this subsection that for $n \geq 1$ and for every language $R \subseteq (A_F)^+$ with $R \in \mathcal{B}_{n+1/2}$ and $R \subseteq WF$ there exists some $T(R) \subseteq A^+$ with $T(R) \in \mathcal{L}_{n+3/2}$ such that $w \in T(R) \Leftrightarrow w_F \in R$ for all $w \in A^+$. We do this by an appropriate transformation of expressions. First, we observe the following two propositions.

Proposition 5.18. *Let $\mu \in WF$. Then $L_\mu \in \mathcal{L}_{3/2}$ and $w \in L_\mu \Leftrightarrow w_F = \mu$ for all $w \in A^+$.*

Proof. Let $\mu = c_1 c_2 \cdots c_n \in WF$ for some $n \geq 1$. Recall that by Definition 5.5 it holds that $L_\mu = \{ v \in A^+ \mid v_F = \mu \}$. So we immediately have $w \in L_\mu \Leftrightarrow w_F = \mu$ for all $w \in A^+$ and it remains to show $L_\mu \in \mathcal{L}_{3/2}$. For $c \in A_F$ define

$$T(c) \quad =_{\text{def}} \quad \begin{cases} a^j & : \quad c = a_j \text{ and } 1 \leq j \leq r-1 \\ b^j & : \quad c = b_j \text{ and } 1 \leq j \leq r-1 \\ a^r \cdot \{a\}^* & : \quad c = a_r \\ b^r \cdot \{b\}^* & : \quad c = b_r \end{cases}$$

and set $T(\mu) =_{\text{def}} T(c_1) \cdot T(c_2) \cdots T(c_n)$. Note with Corollary 2.21.1 that $T(\mu) \in \mathcal{L}_{3/2}$ since we can substitute subsequent letters $d_1 d_2$ with $d_1, d_2 \in A$ by $d_1 \cdot \emptyset^* \cdot d_2$. We show that $L_\mu = T(\mu)$.

Let $w \in A^+$ be given and assume $w \in L_\mu$. Then $w_F = \mu$ and hence w has the A-factorization $w = w_1 w_2 \cdots w_n$ with factors $w_i \in \{a\}^+ \cup \{b\}^+$ of maximal length. By Proposition 5.9 we have $(w_i)_F = c_i$ for $1 \leq i \leq n$. Now fix some i with $1 \leq i \leq n$ and suppose w.l.o.g. that $c_i = a_j$ for some $1 \leq j \leq r$. Then $w_i = a^l$ for some $l \geq 1$ where $j = \min\{r, l\}$. If $1 \leq j \leq r-1$ then $j = l$ and $w_i = a^l = a^j = T(c_i)$. If $j = r$ then $l \geq r$ and $w_i = a^l \in a^r \cdot \{a\}^* = T(c_i)$. We put all factors together and see that $w \in T(\mu)$.

Now suppose $w \in T(\mu)$. Hence, we can write w as $w = v_1 v_2 \cdots v_n$ with $v_i \in T(c_i)$ ($v_i = T(c_i)$, respectively) for $1 \leq i \leq n$. The definition of each $T(c_i)$ is such that $v_i \in T(c_i)$ ($v_i = T(c_i)$, respectively) implies $(v_i)_F = c_i$ for $1 \leq i \leq n$. On the other hand, the types of the c_i alternate between a and b since $\mu \in WF$. So $(v_1)_F (v_2)_F \cdots (v_n)_F \in WF$ and we can apply Proposition 5.8 to obtain $w_F = (v_1 v_2 \cdots v_n)_F = (v_1)_F (v_2)_F \cdots (v_n)_F = c_1 c_2 \cdots c_n = \mu$, i.e., $w \in L_\mu$. ❑

Proposition 5.19. *It holds that $WF \in \mathrm{co}\mathcal{B}_{1/2}$.*

Proof. Consider the following definition.

$$T =_{\mathrm{def}} (A_F)^+ \setminus \left(\bigcup_{\nu \in A_F^a A_F^a \cup A_F^b A_F^b} (A_F)^* \cdot \nu \cdot (A_F)^* \right)$$

Clearly, $T \in \mathrm{co}\mathcal{B}_{1/2}$ and we want to show $T = WF$. Let $\mu = c_1 c_2 \cdots c_n \in (A_F)^+$ for some $n \geq 1$ be given and recall from Proposition 5.6 that $\mu \in WF$ if and only if $c_i \in A_F^a \Leftrightarrow c_{i+1} \in A_F^b$ for $1 \leq i < n$. So if every two consecutive letters in μ have different type, then μ is in none of the sets subtracted from $(A_F)^+$ in the definition of T. Conversely, if two consecutive letters $c_i c_{i+1}$ in μ have the same type, then $c_i c_{i+1} \in A_F^a A_F^a \cup A_F^b A_F^b$ and $\mu \notin T$. ❑

With the following Lemmas 5.20 and 5.21 we prepare the proof of Lemma 5.22. In particular, Lemma 5.21 will serve as a part of the induction base there.

Lemma 5.20. *Let $\mu_0, \mu_1, \ldots, \mu_m \in (A_F)^+$ for some $m \geq 0$ and let $l_1, r_1, l_2, r_2, \ldots, l_m, r_m \in A$ such that $\mu_0 \cdot A_F^{l_1}$, $A_F^{r_i} \cdot \mu_i \cdot A_F^{l_{i+1}}$, $A_F^{r_m} \cdot \mu_m \subseteq WF$ for $1 \leq i < m$. For the language*

$$R = \mu_0 \cdot A_F^{l_1} (A_F)^* A_F^{r_1} \cdot \mu_1 \cdot A_F^{l_2} (A_F)^* A_F^{r_2} \cdot \mu_2 \cdots A_F^{l_m} (A_F)^* A_F^{r_m} \cdot \mu_m$$

there exists some language $T(R) \subseteq A^+$ with $T(R) \in \mathcal{L}_{3/2}$ such that $w_F \in R \Leftrightarrow w \in T(R)$ for all $w \in A^+$.

Proof. We define the transformation of R as

$$T(R) =_{\mathrm{def}} L_{\mu_0} \cdot l_1 B_1 r_1 \cdot L_{\mu_1} \cdot l_2 B_2 r_2 \cdot L_{\mu_2} \cdots l_m B_m r_m \cdot L_{\mu_m}$$

with

$$B_i =_{\mathrm{def}} \begin{cases} A^* & : \; l_i \neq r_i \\ A^* b A^* & : \; l_i = r_i = a \\ A^* a A^* & : \; l_i = r_i = b \end{cases}$$

for $1 \leq i \leq m$. Clearly, $T(R) \subseteq A^+$ and we obtain from Proposition 5.18 that $L_{\mu_i} \in \mathcal{L}_{3/2}$ for all $0 \leq i \leq m$. If we choose a representation for each L_{μ_i} as given by Proposition 2.23.1 we see after distributing the occurring unions over the concatenations that also $T(R)$ has a representation which satisfies Proposition 2.23.1. So $T(R) \in \mathcal{L}_{3/2}$.

Now let $n \geq 1$ and $w = w_1 w_2 \cdots w_n \in A^+$ be the A-factorization of w. Moreover, let $w_F = c_1 c_2 \cdots c_n \in WF$ with $c_i \in A_F$.

Assume first that $w_F \in R$. We can write w_F as

$$w_F = \mu_0 \cdot d_1\nu_1 e_1 \cdot \mu_1 \cdot d_2\nu_2 e_2 \cdot \mu_2 \cdots d_m\nu_m e_m \cdot \mu_m$$

with $d_i \in A_F^{l_i}$, $e_i \in A_F^{r_i}$ and $\nu_i \in (A_F)^*$ for $1 \le i \le m$. Let $1 < j_1 < j'_1 < j_2 < j'_2 \cdots < j_m < j'_m < n$ such that $d_i = c_{j_i}$, $\nu_i = c_{j_i+1} \cdots c_{j'_i-1}$ and $e_i = c_{j'_i}$ for $1 \le i \le m$, and $\mu_i = c_{j'_i+1} \cdots c_{j_{i+1}-1}$ for $0 \le i \le m$ (set $j'_0 =_{\mathrm{def}} 0$ and $j_{m+1} =_{\mathrm{def}} n+1$). We apply Proposition 5.9 and get $\mu_i = (w_{j'_i+1} \cdots w_{j_{i+1}-1})_F$ for $0 \le i \le m$. By Proposition 5.18 this implies that $w_{j'_i+1} \cdots w_{j_{i+1}-1} \in L_{\mu_i}$ for $0 \le i \le m$. Also by Proposition 5.9 we see that $c_{j_i} \cdots c_{j'_i} = d_i\nu_i e_i = (w_{j_i} \cdots w_{j'_i})_F$ with $w_{j_i} \in \{l_i\}^+$ and $w_{j'_i} \in \{r_i\}^+$ for $1 \le i \le m$. Now fix some i with $1 \le i \le m$. If $l_i \ne r_i$ then $w_{j_i} \cdots w_{j'_i} \in l_i A^* r_i = l_i B_i r_i$. If $l_i = r_i$ we may assume w.l.o.g. that $l_i = r_i = a$ and hence $w_{j_i}, w_{j'_i} \in \{a\}^+$. But since the latter are factors of maximal length and because $j_i < j'_i$ there must be some k with $j_i < k < j'_i$ and $w_k \in \{b\}^+$. So $w_{j_i} \cdots w_k \cdots w_{j'_i} \in aA^*bA^*a = l_i B_i r_i$. We put all factors together and see that $w \in T(R)$.

Now assume that $w \in T(R)$ and write w as

$$w = u_0 \cdot l_1 v_1 r_1 \cdot u_1 \cdot l_2 v_2 r_2 \cdot u_2 \cdots l_m v_m r_m \cdot u_m$$

with $v_i \in B_i$ for $1 \le i \le m$ and $u_i \in L_{\mu_i}$ for $0 \le i \le m$. We get from Proposition 5.18 that $(u_i)_F = \mu_i$ for $0 \le i \le m$.

Now fix some i with $1 \le i \le m$ and let $(l_i v_i r_i)_F = f_1 f_2 \cdots f_{n'}$ for some $n' \ge 1$ and $f_j \in A_F$. First suppose $l_i \ne r_i$. Then f_1 has type l_i and $f_{n'}$ has different type r_i. So $n' \ge 2$ and $f_1 f_2 \cdots f_{n'} \in A_F^{l_i}(A_F)^* A_F^{r_i}$. If $l_i = r_i$ we may assume w.l.o.g. that $l_i = r_i = a$. Then f_1 and $f_{n'}$ have both type a. By definition of $T(R)$ we get $v_i \in B_i = A^*bA^*$. It follows that there is some factor from $\{b\}^+$ in av_ia and hence, there must be some k with $1 < k < n'$ such that f_k has type b. In particular, we see that $n' \ge 3$ and so $f_1 f_2 \cdots f_{n'} \in A_F^a (A_F)^* A_F^a = A_F^{l_i}(A_F)^* A_F^{r_i}$.

Define $\varphi_i =_{\mathrm{def}} (l_i v_i r_i)_F$ for $1 \le i \le m$. We have just shown that $\varphi_i \in A_F^{l_i}(A_F)^* A_F^{r_i}$ for $1 \le i \le m$ and obtain

$$\mu \;=_{\mathrm{def}}\; (u_0)_F \cdot (l_1 v_1 r_1)_F \cdot (u_1)_F \cdots (l_m v_m r_m)_F \cdot (u_m)_F \;=\; \mu_0 \cdot \varphi_1 \cdot \mu_1 \cdots \varphi_m \cdot \mu_m \;\in\; R.$$

Now we argue that μ is well–formed for which it is sufficient to observe that $l(\mu_0) \ne f(\varphi_1) = l_1$, that $r_m = l(\varphi_m) \ne f(\mu_m)$, and that for $1 \le i < m$ it holds that $r_i = l(\varphi_i) \ne f(\mu_i)$ and $l(\mu_i) \ne f(\varphi_{i+1}) = l_{i+1}$. But all this is clear since we have assumed in the lemma that $\mu_0 \cdot A_F^{l_1}$, $A_F^{r_i} \cdot \mu_i \cdot A_F^{l_{i+1}}$, $A_F^{r_m} \cdot \mu_m \in WF$ for $1 \le i < m$.

Finally, we apply repeatedly Proposition 5.7 and Proposition 5.8 to $\mu \in WF$ and obtain

$$(u_0)_F \cdot (l_1 v_1 r_1)_F \cdot (u_1)_F \cdots (l_m v_m r_m)_F \cdot (u_m)_F \;=\; (u_0 \cdot l_1 v_1 r_1 \cdot u_1 \cdots l_m v_m r_m \cdot u_m)_F.$$

So $w_F = \mu \in R$. ❑

Let $B \subseteq A^+$, $C \subseteq (A_F)^+$. We write for short $\overline{B}$ instead of $A^+ \setminus B$ and $\overline{C}$ instead of $(A_F)^+ \setminus C$.

Lemma 5.21. *Let $R \subseteq (A_F)^+$ with $R \subseteq WF$ and $R \in \mathrm{co}\mathcal{B}_{1/2}$. Then there exists some language $T(R) \subseteq A^+$ with $T(R) \in \mathrm{co}\mathcal{L}_{3/2}$ such that $w_F \in R \Leftrightarrow w \in T(R)$ for all $w \in A^+$.*

Proof. By assumption we have $\overline{R} \in \mathcal{B}_{1/2}$, hence we can write $\overline{R}$ as a finite union of languages R' of the form

$$R' = \mu_0(A_F)^+\mu_1(A_F)^+ \cdots \mu_{m-1}(A_F)^+\mu_m$$

for $m \ge 0$ and $\mu_i \in (A_F)^*$ for $0 \le i \le m$ (see Lemma 2.17). Let M denote the finite union of all letters in A_F. We can assume w.l.o.g. that for $0 \le i \le m$ we have $\mu_i \ne \varepsilon$. To see this assume first that $\mu_0 = \varepsilon$. Then it must be that $m \ge 1$ and we rewrite the first $(A_F)^+$ as $M \cup M(A_F)^+$, and distribute the occurring unions in M over the concatenation. If $\mu_m = \varepsilon$ we can do a similar thing. Finally, if $\mu_i = \varepsilon$ for $0 < i < m$ we write $(A_F)^+(A_F)^+$ as $M(A_F)^+$ and continue as before.

Moreover, we can write each $(A_F)^+$ as $M \cup A_F(A_F)^*A_F$ and again, distribute the occurring unions in M over the concatenation. So we get as a first step that $\overline{R}$ is a finite union of languages R' of the form

$$R' = \mu_0 \cdot (A_F^a \cup A_F^b) \cdot (A_F)^* \cdot (A_F^a \cup A_F^b) \cdot \mu_1 \cdots (A_F^a \cup A_F^b) \cdot (A_F)^* \cdot (A_F^a \cup A_F^b) \cdot \mu_m$$

for some $m \geq 0$ and $\mu_i \in (A_F)^+$ for $0 \leq i \leq m$. Now we distribute the remaining unions $(A_F^a \cup A_F^b)$ and find that $\overline{R}$ is a finite union of languages R' of the form

$$R' = \mu_0 \cdot A_F^{l_1} \cdot (A_F)^* \cdot A_F^{r_1} \cdot \mu_1 \cdot A_F^{l_2} \cdot (A_F)^* \cdot A_F^{r_2} \cdot \mu_2 \cdots A_F^{l_m} \cdot (A_F)^* \cdot A_F^{r_m} \cdot \mu_m \tag{20}$$

for some $m \geq 0$, letters $l_1, r_1, l_2, r_2, \ldots, l_m, r_m \in A$ and $\mu_i \in (A_F)^+$ for $0 \leq i \leq m$.

Observe that if for some R' the condition

$$\left(\mu_0 A_F^{l_1} \quad \cup \quad \bigcup_{1 \leq i < m} A_F^{r_i} \mu_i A_F^{l_{i+1}} \quad \cup \quad A_F^{r_m} \mu_m \right) \quad \cap \quad \overline{WF} \quad \neq \quad \emptyset \tag{21}$$

is true, then $R' \subseteq \overline{WF}$, which can be seen as follows. Suppose ν witnesses that (21) holds and let us further assume w.l.o.g. that $\nu \in A_F^{r_i} \mu_i A_F^{l_{i+1}}$ for some fixed i with $1 \leq i < m$. (the cases $\nu \in \mu_0 A_F^{l_1}$ and $\nu \in A_F^{r_m} \mu_m$ are completely analogous). Then also any other $\nu' \in A_F^{r_i} \mu_i A_F^{l_{i+1}}$ is in $\overline{WF}$ since ν and ν' have the same sequence of types of letters from A_F. Now if there is some $\beta \in R' \cap WF$ then we have by definition of R' that there is some $\nu' \in A_F^{r_i} \mu_i A_F^{l_{i+1}}$ such that $\nu' \in \overline{WF}$ is a factor of $\beta \in WF$, contradicting Proposition 5.7.

Suppose that for $k \geq 0$ the sets $R'_1, R'_2, \ldots, R'_k$ of the form as stated in (20) occur in the finite union describing $\overline{R}$. We turn to the description of R now. Let $I \subseteq \{1, \ldots, k\}$ be the set of indices such that R'_i does not satisfy (21) for all $i \in I$ and set $\overline{I} =_{\text{def}} \{1, \ldots, k\} \backslash I$. As pointed out before, we have $WF \subseteq \overline{R'_i}$ for all $i \in \overline{I}$. Recall that $R \subseteq WF$ is a prerequisite of this lemma. So we have

$$\begin{aligned}
R \quad &= \quad R \ \cap \ WF && \text{since } R \subseteq WF \\
&= \quad \textstyle\bigcap_{1 \leq i \leq k} \overline{R'_i} \ \cap \ WF \\
&= \quad \textstyle\bigcap_{i \in I} \overline{R'_i} \ \cap \ \bigcap_{i \in \overline{I}} \overline{R'_i} \ \cap \ WF \\
&= \quad \textstyle\bigcap_{i \in I} \overline{R'_i} \ \cap \ WF && \text{because } WF \ \subseteq \ \textstyle\bigcap_{i \in \overline{I}} \overline{R'_i}.
\end{aligned}$$

Note here that if we only intersect each R'_i with WF instead of making the distinction with condition (21), then we can not show that the resulting language is in $\mathcal{B}_{1/2}$. Since for all $i \in I$ the sets R'_i do not fulfil (21) it follows that we can apply Lemma 5.20 to each of them. Denote for all $i \in I$ by $T(R'_i)$ the languages of A^+ with $T(R'_i) \in \mathcal{L}_{3/2}$ provided by Lemma 5.20 and define

$$T(R) \quad =_{\text{def}} \quad \bigcap_{i \in I} \overline{T(R'_i)} \, .$$

The class $\text{co}\mathcal{L}_{3/2}$ is closed under intersection by definition, so $T(R) \in \text{co}\mathcal{L}_{3/2}$. Let $w \in A^+$ be given. One easily verifies that

$$\begin{aligned}
w_F \in R \quad &\Longleftrightarrow \quad w_F \in WF \text{ and } w_F \in \overline{R'_i} \text{ for all } i \in I \\
&\Longleftrightarrow \quad w_F \in \overline{R'_i} \text{ for all } i \in I && \text{since } w_F \in WF \\
&\Longleftrightarrow \quad w \in \overline{T(R'_i)} \text{ for all } i \in I && \text{since } w_F \in R'_i \Leftrightarrow w \in T(R'_i) \text{ by 5.20} \\
&\Longleftrightarrow \quad w \in T(R).
\end{aligned}$$

❑

Lemma 5.22. *Let $R \subseteq (A_F)^+$ with $R \subseteq WF$ and let $n \geq 1$.*

1. *If $R \in \mathrm{co}\mathcal{B}_{n-1/2}$ then there exists some language $T(R) \subseteq A^+$ with $T(R) \in \mathrm{co}\mathcal{L}_{n+1/2}$ such that $w_F \in R \Leftrightarrow w \in T(R)$ for all $w \in A^+$.*
2. *If $R \in \mathcal{B}_{n+1/2}$ then there exists some language $T(R) \subseteq A^+$ with $T(R) \in \mathcal{L}_{n+3/2}$ such that $w_F \in R \Leftrightarrow w \in T(R)$ for all $w \in A^+$.*

Proof. We prove both statements simultaneously by induction on n.
Induction base. Let $n = 1$. Then statement 1 holds by Lemma 5.21 and we have to show statement 2. Recall from Lemma 2.24.2 that $\mathcal{B}_{3/2} = \mathrm{Pol}(\mathrm{co}\mathcal{B}_{1/2})$. So R can be written as a finite union of languages R' for which in turn there are languages $L_0, L_1, \ldots, L_m \subseteq (A_F)^+$ for some $m \geq 0$ such that

$$R' = L_0 L_1 \cdots L_m$$

with $L_i \in \mathrm{co}\mathcal{B}_{1/2}$ for $0 \leq i \leq m$. From $R \subseteq WF$ we get $R' \subseteq WF$ for each member R' of the union. With Proposition 5.7 we see that for each R' it holds that $L_i \subseteq WF$ for all $0 \leq i \leq m$. Let us define the transformation $T(L_i) \subseteq A^+$ of L_i to be the set from $\mathrm{co}\mathcal{L}_{3/2}$ provided by statement 1. Now set

$$T(R') =_{\mathrm{def}} T(L_0)T(L_1)\cdots T(L_m)$$

and define $T(R)$ to be the union of all $T(R')$ going over all occurring R'.

Clearly, $T(R) \subseteq A^+$ since $T(L_i) \subseteq A^+$. Let us verify that $T(R) \in \mathcal{L}_{5/2}$. Since $T(L_i) \in \mathrm{co}\mathcal{L}_{3/2}$ for $0 \leq i \leq m$ we have $T(R') \in \mathrm{Pol}(\mathrm{co}\mathcal{L}_{3/2})$ and also $T(R) \in \mathrm{Pol}(\mathrm{co}\mathcal{L}_{3/2})$. So $T(R) \in \mathcal{L}_{5/2}$ by Lemma 2.24.1.

It remains to show that $w_F \in R \Leftrightarrow w \in T(R)$ for all $w \in A^+$. We can restrict ourselves to a single member of the union and show $w_F \in R' \Leftrightarrow w \in T(R')$ for all $w \in A^+$. Let $n \geq 1$ and $w = w_1 w_2 \cdots w_n \in A^+$ be the A-factorization of w. Moreover, let $w_F = c_1 c_2 \cdots c_n \in WF$ with $c_i \in A_F$.

Assume first that $w_F \in R'$. We can write w_F as $w_F = \mu_0 \mu_1 \cdots \mu_m$ with $\mu_i \in L_i$ for $0 \leq i \leq m$. Let $1 = j_0 < j_1 < \cdots < j_m < n$ such that $\mu_i = c_{j_i} \cdots c_{j_{i+1}-1}$ for $0 \leq i \leq m$ (set $j_{m+1} =_{\mathrm{def}} n+1$). We can apply Proposition 5.9 and get $\mu_i = (w_{j_i} \cdots w_{j_{i+1}-1})_F$ for $0 \leq i \leq m$. From statement 1 we obtain that $(w_{j_i} \cdots w_{j_{i+1}-1})_F \in L_i$ implies $w_{j_i} \cdots w_{j_{i+1}-1} \in T(L_i)$ for $0 \leq i \leq m$. If we put these pieces together we get $w \in T(R')$.

Now assume conversely that $w \in T(R')$. Let $w = u_0 u_1 \cdots u_m$ with $u_i \in T(L_i)$ for $0 \leq i \leq m$. We get from statement 1 that $(u_i)_F \in L_i$ for $0 \leq i \leq m$. So $(u_0)_F (u_1)_F \cdots (u_m)_F \in R' \subseteq WF$. We apply Proposition 5.7 and Proposition 5.8 to see that $(u_0)_F (u_1)_F \cdots (u_m)_F = (u_0 u_1 \cdots u_m)_F = w_F \in R'$.

Induction step. Assume the lemma holds for some $n \geq 1$ and we want to show that it also holds for $n+1$. First suppose that $R \in \mathrm{co}\mathcal{B}_{n+1/2}$ and define $R' =_{\mathrm{def}} \overline{R} \cap WF$. Since $n \geq 1$ and because $WF \in \mathrm{co}\mathcal{B}_{1/2} \subseteq \mathcal{B}_{n+1/2}$ by Proposition 5.19, we obtain from the closure of $\mathcal{B}_{n+1/2}$ under intersection (see Lemma 2.25) that $R' \in \mathcal{B}_{n+1/2}$. Moreover, we have

$$\overline{R'} \cap WF \quad = \quad \overline{\overline{R} \cap WF} \cap WF \quad = \quad \left(R \cup \overline{WF}\right) \cap WF \quad = \quad R \cap WF. \tag{22}$$

So we see that

$$\begin{aligned} R &= R \cap WF && \text{since } R \subseteq WF \\ &= \overline{R'} \cap WF && \text{because of Eq. (22).} \end{aligned}$$

Since R' is a subset of WF, we can apply the induction hypothesis of statement 2. Denote by $T(R')$ the language of A^+ with $T(R') \in \mathcal{L}_{n+3/2}$ provided by the hypothesis and define

$$T(R) \quad =_{\mathrm{def}} \quad \overline{T(R')}\,.$$

Note that $T(R) \in \mathrm{co}\mathcal{L}_{n+3/2}$. Let $w \in A^+$ be given. One easily verifies that

$$\begin{aligned} w_F \in R &\iff w_F \in WF \text{ and } w_F \in \overline{R'} \\ &\iff w_F \in \overline{R'} && \text{since } w_F \in WF \\ &\iff w \in \overline{T(R')} && \text{since } w_F \in R' \Leftrightarrow w \in T(R') \text{ by hypothesis} \\ &\iff w \in T(R). \end{aligned}$$

This completes the induction step for statement 1 and we turn to statement 2.

Let $R \in \mathcal{B}_{n+3/2}$ with $R \subseteq WF$ be given and recall from Lemma 2.24.2 that $\mathcal{B}_{n+3/2} = \mathrm{Pol}(\mathrm{co}\mathcal{B}_{n+1/2})$. Now we can proceed exactly as in the proof of the induction base for statement 2. Only observe that we can use what we have just shown for $\mathrm{co}\mathcal{B}_{n+1/2}$ in the induction step for statement 1. So there exists some language $T(R) \subseteq A^+$ with $T(R) \in \mathcal{L}_{n+5/2}$ such that $w_F \in R \Leftrightarrow w \in T(R)$ for all $w \in A^+$. ❑

5.5 Connecting Hierarchies

Recall that we started with $L = L(F) \subseteq A^+$ and obtained $\widetilde{L} \subseteq (A_F)^+$ via the construction of $\widetilde{F}$. We now apply Lemma 5.22 to $\widetilde{L}$.

Lemma 5.23. *Let $n \geq 1$. If $\widetilde{L} \in \mathcal{B}_{n+1/2}$ then $L \in \mathcal{L}_{n+3/2}$.*

Proof. By Proposition 5.13 we have $\widetilde{L} = \{ w_F \mid w \in L \}$, so in particular $\widetilde{L} \subseteq WF$. Since we assume $\widetilde{L} \in \mathcal{B}_{n+1/2}$, we get from Lemma 5.22 a language $T(\widetilde{L}) \subseteq A^+$ with $T(\widetilde{L}) \in \mathcal{L}_{n+3/2}$. To show $T(\widetilde{L}) = L$ let $w \in A^+$ be given. Now we have

$$\begin{aligned} w \in L &\iff w_F \in \widetilde{L} && \text{by Proposition 5.13} \\ &\iff w \in T(\widetilde{L}) && \text{by Lemma 5.22.} \end{aligned}$$

❑

As the main result of this section we have the following theorem.

Theorem 5.24. *Let $n \geq 1$. Suppose it holds that $\mathcal{C}_n^{\mathcal{B}} = \mathcal{B}_{n+1/2}$ for arbitrary alphabets. Then $\mathcal{C}_{n+1}^{\mathcal{L}} = \mathcal{L}_{n+3/2}$ for a two–letter alphabet.*

Proof. The inclusion $\mathcal{L}_{n+3/2} \subseteq \mathcal{C}_{n+1}^{\mathcal{L}}$ is from Theorem 4.6 and holds unconditionally. For the other inclusion let $L \in \mathcal{C}_{n+1}^{\mathcal{L}}$. By Theorem 4.7 and Theorem 2.1 there is some minimal permutationfree dfa F with $L = L(F)$ and which does not have pattern $\mathbb{L}_{n+1}^{\mathcal{L}}$. So we can do the construction of $\widetilde{F}$ and we apply Corollary 5.17 to obtain $\widetilde{L} \in \mathcal{C}_n^{\mathcal{B}}$. By assumption, $\mathcal{C}_n^{\mathcal{B}} = \mathcal{B}_{n+1/2}$. We apply Lemma 5.23 to see $L \in \mathcal{L}_{n+3/2}$. ❑

From Theorem 4.6 we know that in particular $\mathcal{C}_1^{\mathcal{B}} = \mathcal{B}_{3/2}$ for arbitrary alphabets.

Corollary 5.25. *Let $|A| = 2$. Then $\mathcal{C}_2^{\mathcal{L}} = \mathcal{L}_{5/2}$.*

The following is an immediate consequence of Theorem 4.15.

Corollary 5.26. *Let $|A| = 2$ and $L \subseteq A^+$. The question $L \in \mathcal{L}_{5/2}$ is decidable.*

References

[Arf87] M. Arfi. Polynomial operations on rational languages. In *Proceedings 4th Symposium on Theoretical Aspects of Computer Science*, volume 247 of *Lecture Notes in Computer Science*, pages 198–206. Springer-Verlag, 1987.

[Arf91] M. Arfi. Opérations polynomiales et hiérarchies de concaténation. *Theoretical Computer Science*, 91:71–84, 1991.

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

[CB71] R. S. Cohen and J. A. Brzozowski. Dot-depth of star-free events. *Journal of Computer and System Sciences*, 5:1–16, 1971.

[CH91] S. Cho and D. T. Huynh. Finite-automaton aperiodicity is PSPACE-complete. *Theoretical Computer Science*, 88:99–116, 1991.

[Eil76] S. Eilenberg. *Automata, languages and machines*, volume B. Academic Press, New York, 1976.

[Gla98] C. Glaßer. A normalform for classes of concatenation hierarchies. Technical Report 216, Inst. für Informatik, Univ. Würzburg, 1998.

[GS99] C. Glaßer and H. Schmitz. Languages of dot-depth 3/2. Technical Report 243, Inst. für Informatik, Univ. Würzburg, 1999. Available via ftp from *http://www.informatik.uni-wuerzburg.de/reports/tr.html*.

[GS00] C. Glaßer and H. Schmitz. Languages of dot-depth 3/2. In *Proceedings 17th Symposium on Theoretical Aspects of Computer Science*, volume 1770 of *LNCS*, pages 555–566. Springer Verlag, 2000.

[Imm88] N. Immerman. Nondeterministic space is closed under complementation. *SIAM Journal on Computing*, 17:935–938, 1988.

[Kna83] R. Knast. A semigroup characterization of dot-depth one languages. *RAIRO Inform. Théor.*, 17:321–330, 1983.

[MP71] R. McNaughton and S. Papert. *Counterfree Automata*. MIT Press, Cambridge, 1971.

[Pin95] J.-E. Pin. Finite semigroups and recognizable languages: an introduction. In J. Fountain, editor, *NATO Advanced Study Institute: Semigroups, Formal Languages and Groups*, pages 1–32. Kluwer Academic Publishers, 1995.

[Pin96] J.-E. Pin. Syntactic semigroups. In G.Rozenberg and A.Salomaa, editors, *Handbook of formal languages*, volume I, pages 679–746. Springer, 1996.

[Pin98] J.-E. Pin. Bridges for concatenation hierarchies. In *Proceedings 25th ICALP*, volume 1443 of *LNCS*, pages 431–442. Springer Verlag, 1998.

[PP86] D. Perrin and J.-E. Pin. First-order logic and star-free sets. *Journal of Computer and System Sciences*, 32:393–406, 1986.

[PW97] J.-E. Pin and P. Weil. Polynomial closure and unambiguous product. *Theory of computing systems*, 30:383–422, 1997.

[RST82] W. L. Ruzzo, J. Simon, and M. Tompa. Space-bounded hierarchies and probabilistic computations. In *Proceedings of the 14th ACM Symposium on Theory of Computing*, pages 215–223, San Francisco, California, 5–7 May 1982.

[Sch65] M. P. Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.

[Sim75] I. Simon. Piecewise testable events. In *Proceedings 2nd GI Conference*, volume 33 of *Lecture Notes in Computer Science*, pages 214–222. Springer-Verlag, 1975.

[Str81] H. Straubing. A generalization of the Schützenberger product of finite monoids. *Theoretical Computer Science*, 13:137–150, 1981.

[Str85] H. Straubing. Finite semigroup varieties of the form V * D. *J.Pure Appl.Algebra*, 36:53–94, 1985.

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

[SW98] H. Schmitz and K. W. Wagner. The Boolean hierarchy over level 1/2 of the Straubing-Thèrien hierarchy. Technical Report 201, Inst. für Informatik, Univ. Würzburg, 1998.

[Sze87] R. Szelepcsényi. The method of forcing for nondeterministic automata. *Bull.of the EATCS*, 33:96–100, 1987.

[Thé81] D. Thérien. Classification of finite monoids: the language approach. *Theoretical Computer Science*, 14:195–208, 1981.

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

[Tho84] W. Thomas. An application of the Ehrenfeucht–Fraïssé game in formal language theory. *Société Mathématique de France, mémoire 16*, 2:11–21, 1984.

[Wei93] P. Weil. Some results on the dot-depth hierarchy. *Semigroup Forum*, 46:352–370, 1993.

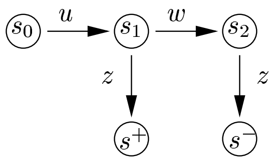


Figure 4: Forbidden Pattern for $\mathcal{L}_{1/2}$. Note that $\mathcal{L}_{1/2} = \mathcal{C}_0^{\mathcal{L}}$, and $p = (\varepsilon, w) \in \mathbb{L}_0^{\mathcal{L}}$.

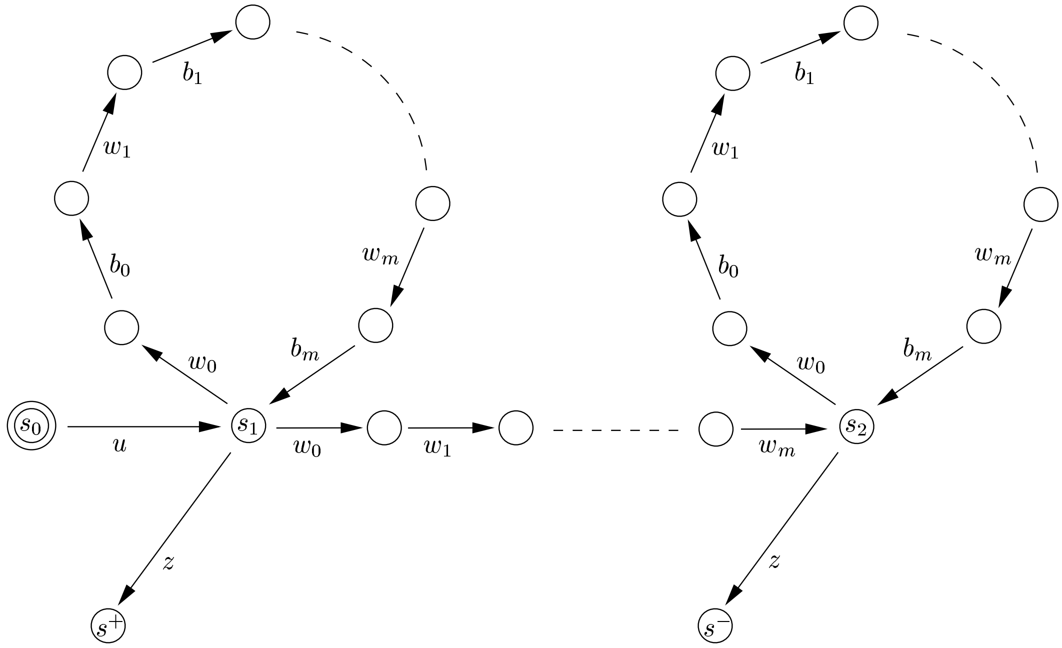


Figure 5: Forbidden Pattern for $\mathcal{L}_{3/2}$. Note that $\mathcal{L}_{3/2} = \mathcal{C}_1^{\mathcal{L}}$, and $w_i \in A^+$, $p_i = (\varepsilon, b_i) \in \mathbb{L}_0^{\mathcal{L}}$, and $p = (w_0, p_0, \ldots, w_m, p_m) \in \mathbb{L}_1^{\mathcal{L}}$.

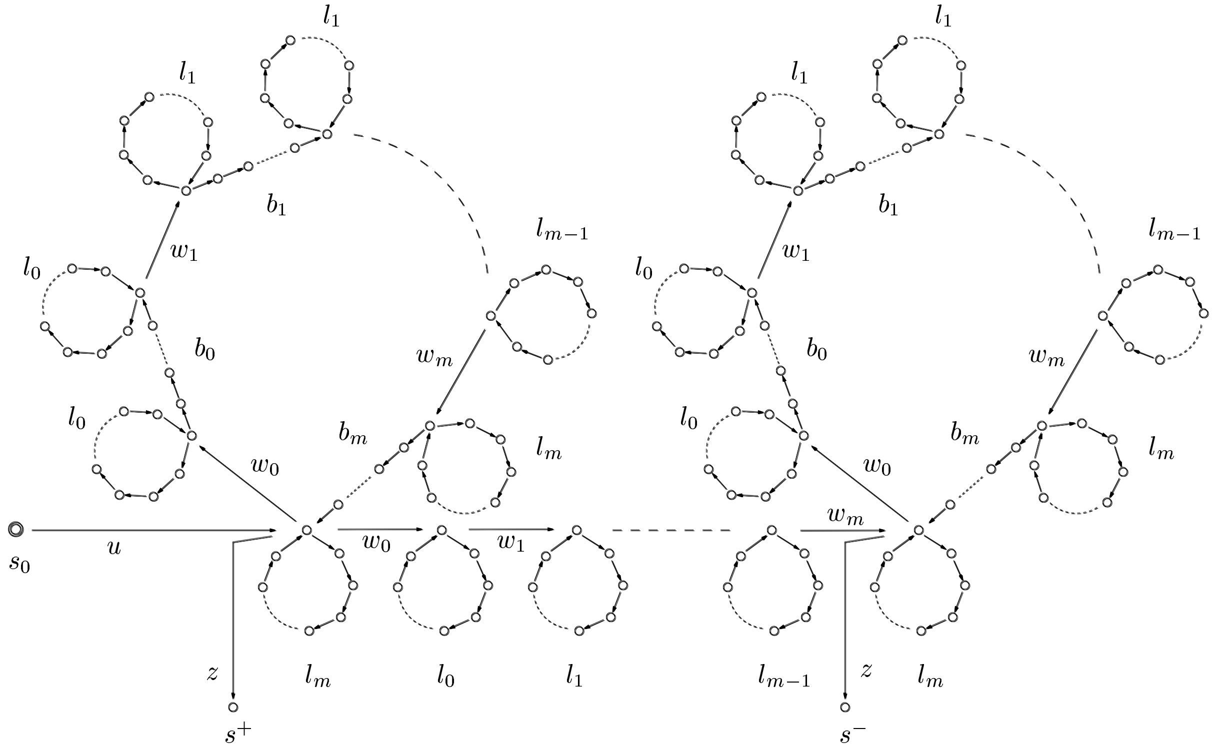


Figure 6: Forbidden Pattern for $\mathcal{C}_2^{\mathcal{L}}$. Note that $\mathcal{L}_{5/2} = \mathcal{C}_2^{\mathcal{L}}$ if $|A| = 2$, and $w_i \in A^+$, $p_i \in \mathbb{L}_1^{\mathcal{L}}$ with $\overline{p_i} = b_i$ and $\overline{p_i}^{\circ} = l_i$, and $p = (w_0, p_0, \ldots, w_m, p_m) \in \mathbb{L}_2^{\mathcal{L}}$.

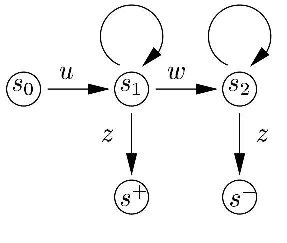


Figure 7: Forbidden Pattern for $\mathcal{B}_{1/2}$. Note that $\mathcal{B}_{1/2} = \mathcal{C}_0^{\mathcal{B}}$, and $p = (v, w) \in \mathbb{L}_0^{\mathcal{B}}$.

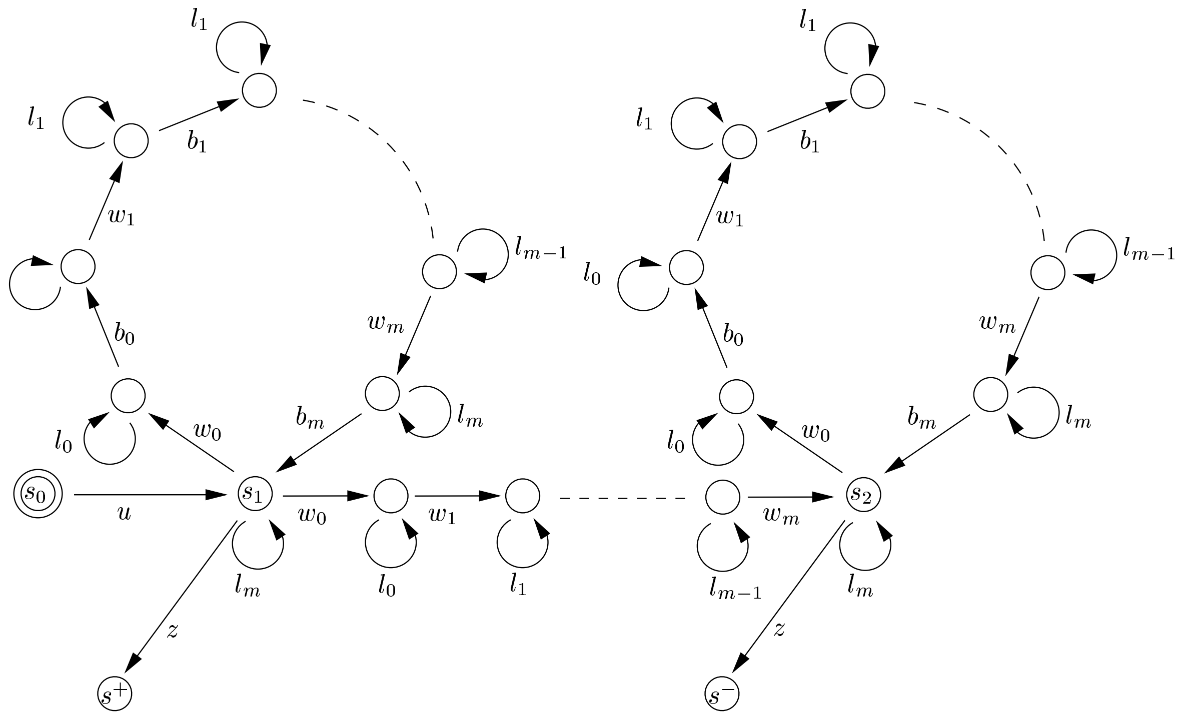


Figure 8: Forbidden Pattern for $\mathcal{B}_{3/2}$. Note that $\mathcal{B}_{3/2} = \mathcal{C}_1^{\mathcal{B}}$, and $w_i \in A^+$, $p_i = (l_i, b_i) \in \mathbb{L}_0^{\mathcal{B}}$, and $p = (w_0, p_0, \ldots, w_m, p_m) \in \mathbb{L}_1^{\mathcal{B}}$.

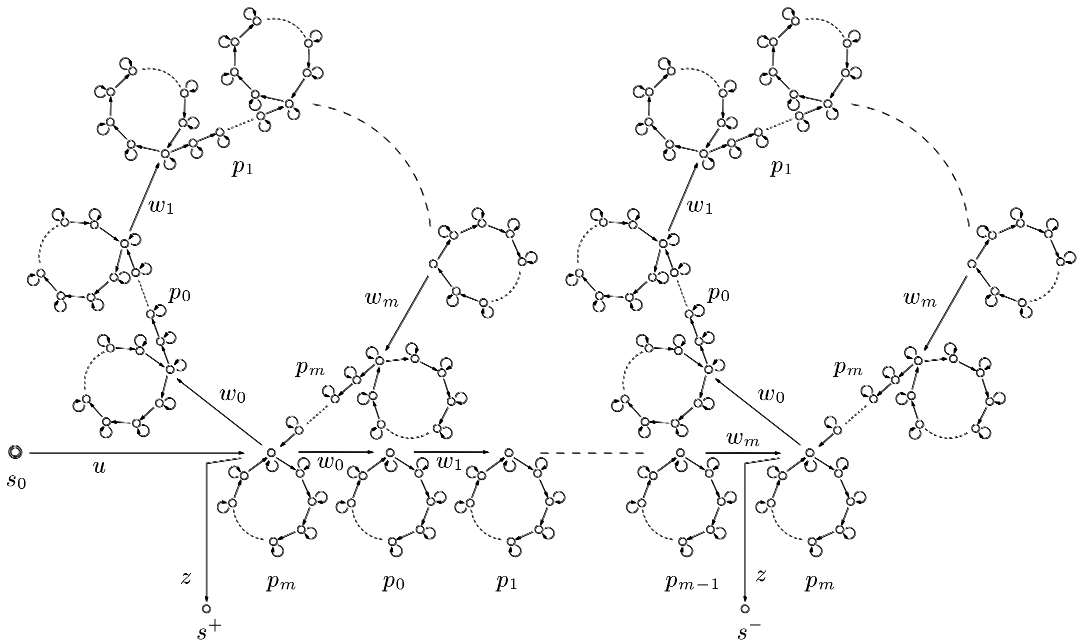


Figure 9: Forbidden Pattern for $\mathcal{C}_2^{\mathcal{B}}$. Note that $w_i \in A^+$, $p_i \in \mathbb{L}_1^{\mathcal{B}}$, and $p = (w_0, p_0, \ldots, w_m, p_m) \in \mathbb{L}_2^{\mathcal{B}}$.

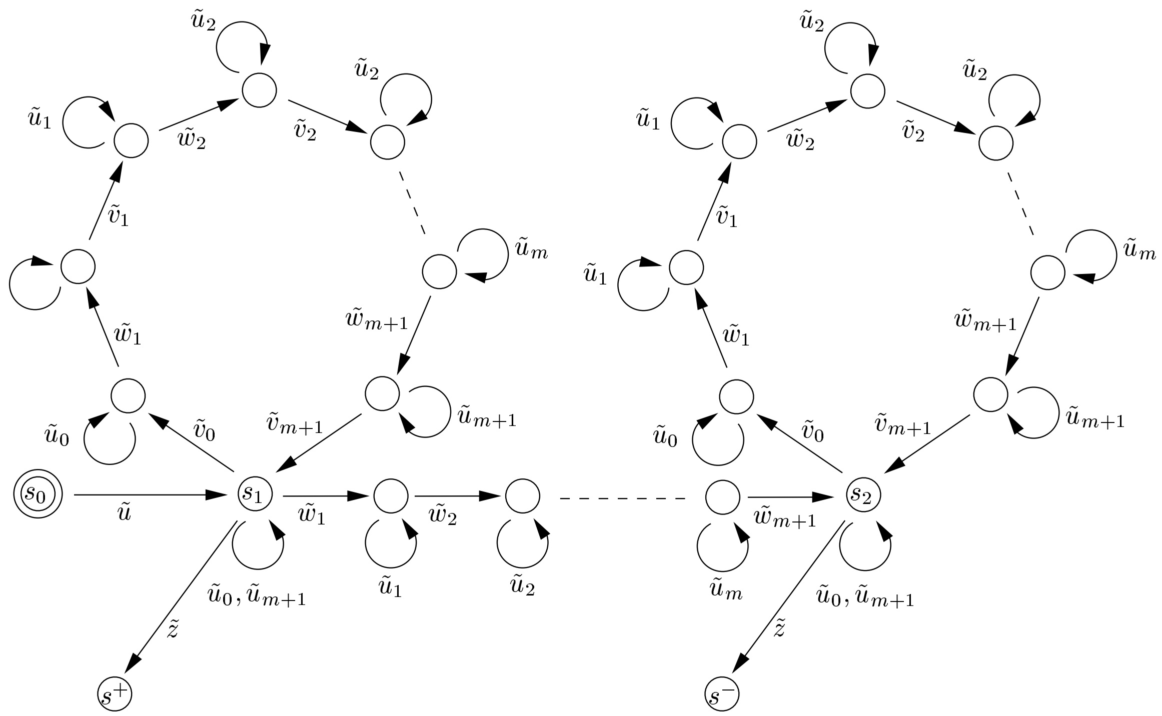


Figure 10: Forbidden Pattern for $\mathcal{B}_{3/2}$ from [GS00], with accepting state s^+, rejecting state s^-, $m \geq 0$, $\tilde{u_i}, \tilde{w_i} \in A^+$ and $\tilde{u}, \tilde{z}, \tilde{v_i} \in A^*$.