

# Forbidden-Patterns and Word Extensions for Concatenation Hierarchies

Dissertation zur Erlangung des  
naturwissenschaftlichen Doktorgrades  
der Bayerischen Julius-Maximilians-Universität Würzburg

vorgelegt von

Christian Glaßer

aus Jena

Würzburg, 2001

Eingereicht am: 23.04.2001  
bei der Fakultät für Mathematik und Informatik  
1. Gutachter: Prof. Dr. Klaus W. Wagner  
2. Gutachter: Prof. Dr. Jean-Eric Pin  
Tag der mündlichen Prüfung: 07.09.2001

Ich danke meinen Eltern und meiner Großmutter, die mir schon früh die Chance gaben, eigene Entscheidungen zu treffen, und auf deren Unterstützung ich mich stets verlassen konnte. Ebenso bin ich meiner Freundin Iljana sehr dankbar, da sie mir des Öfteren zeigte, dass es neben der Informatik noch andere interessante Dinge gibt (z.B. Mathematik ☺). Ohne Iljana wären Studium und Promotionszeit bei Weitem nicht so schön und abwechslungsreich gewesen.

Ich danke allen, die mein Interesse an der wissenschaftlichen Arbeit und insbesondere der theoretischen Informatik geweckt haben. An dieser Stelle möchte ich besonders meine Lehrer Klaus Haberland und Gerd Wechsung hervorheben, deren enthusiastische Vorlesungen Highlights meines Studiums waren und mich schließlich für die Forschung begeisterten.

Ebenfalls möchte ich allen denjenigen danken, die durch Rat und Diskussionen zum Entstehen dieser Arbeit beigetragen haben. Besonders danke ich meinem Doktorvater Klaus W. Wagner und meinem Kollegen Heinz Schmitz für die hervorragende Zusammenarbeit.

Weiterhin möchte ich mich bei Nortel External Research und der Studienstiftung des Deutschen Volkes für die finanzielle Unterstützung im Rahmen zweier Stipendien bedanken.



# Table of Contents

<b>Introduction</b> .....	7
<b>1. Preliminaries</b> .....	13
1.1 Definitions and Notations .....	13
1.2 Well Partial Ordered Sets .....	16
1.3 The Loop-Lemma .....	17
1.4 $\langle_v^{0,k}$ and $\langle_v^{1,k}$ Word Extensions .....	21
1.4.1 Basic Definitions .....	22
1.4.2 The Classes $\tilde{\mathcal{B}}_{k,m}$ .....	23
1.4.3 The $\langle_v^{1,k}$ Upward Closure of a Word .....	24
1.4.4 The $\langle_v^{1,k}$ Upward Closure of Languages from $\tilde{\mathcal{B}}_{k,1}$ .....	28
1.4.5 The $\langle_v^{1,k}$ Upward Closure of Languages from $\tilde{\mathcal{B}}_k$ .....	29
<b>2. Concatenation Hierarchies</b> .....	33
2.1 Definitions of Concatenation Hierarchies .....	33
2.2 Alternative Definitions .....	36
2.3 Normalforms and Closure Properties .....	37
2.4 The Dot-Depth Problem .....	37
2.5 Known Forbidden-Pattern Characterizations .....	38
<b>3. Forbidden-Pattern Theory</b> .....	41
3.1 Looking for an Iteration Rule .....	42
3.2 Forbidden-Pattern Hierarchies .....	43
3.2.1 How to Define Forbidden-Pattern Hierarchies .....	44
3.2.2 Transformations of Patterns .....	46
3.3 Pattern Iterator versus Polynomial Closure .....	51
3.4 Inclusion Structure of the Forbidden-Pattern Classes $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$ .....	54
3.5 Pattern Iteration remains Starfree .....	56
3.6 Decidability of the Forbidden-Pattern Classes .....	62
3.7 Consequences for Concatenation Hierarchies .....	67
3.7.1 Inclusion Relations between DDH, STH and Forbidden-Pattern Classes .....	68
3.7.2 The Forbidden-Pattern Classes are Starfree .....	70
3.7.3 The Hierarchies of Forbidden-Pattern Classes are Strict .....	71
3.7.4 The Classes $\mathcal{FP}(\mathbb{P}_n^{\mathcal{E}})$ and $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$ are Decidable .....	74
3.7.5 Lower Bounds for the Dot-Depth Problem .....	75

3.8	Summary and Discussion . . . . .	75
<b>4.</b>	<b>Decidability Results for the DDH and STH . . . . .</b>	<b>77</b>
4.1	The Levels $1/2$ . . . . .	78
4.1.1	Definition of $\preceq_{\mathcal{M}}^{0,k}$ Word Extensions . . . . .	78
4.1.2	The $\preceq_{\mathcal{M}}^{0,k}$ Upward Closure of a Word . . . . .	80
4.1.3	$\preceq_{\mathcal{M}}^{0,k}$ Co-Ideals are Finitely Generated . . . . .	83
4.1.4	Languages from $\mathcal{FP}(\mathbb{P}_0^{\mathcal{L}})$ and $\mathcal{FP}(\mathbb{P}_0^{\mathcal{B}})$ are $\preceq_{\mathcal{M}}^{0,k}$ Co-Ideals . . . . .	85
4.1.5	$\mathcal{L}_{1/2}$ and $\mathcal{B}_{1/2}$ are decidable . . . . .	86
4.2	The Boolean Hierarchies over the Levels $1/2$ . . . . .	87
4.2.1	$m_{\mathcal{M}}^{0,k}$ characterizes the Boolean Hierarchies over the Levels $1/2$ . . . . .	88
4.2.2	$m_{\mathcal{M}}^{0,k}$ is computable . . . . .	92
4.3	The Levels $3/2$ . . . . .	98
4.3.1	Definition of $\preceq_{\mathcal{M}}^{1,k}$ Word Extensions . . . . .	99
4.3.2	The $\preceq_{\mathcal{M}}^{1,k}$ Upward Closure of a Word . . . . .	103
4.3.3	Regular $\preceq_{\mathcal{M}}^{1,k}$ Co-Ideals are Finitely Generated . . . . .	109
4.3.4	Languages from $\mathcal{FP}(\mathbb{P}_1^{\mathcal{L}})$ and $\mathcal{FP}(\mathbb{P}_1^{\mathcal{B}})$ are Regular $\preceq_{\mathcal{M}}^{1,k}$ Co-Ideals . . . . .	114
4.3.5	$\mathcal{L}_{3/2}$ and $\mathcal{B}_{3/2}$ are decidable . . . . .	117
4.4	Summary and Discussion . . . . .	119
	<b>Bibliography . . . . .</b>	<b>121</b>
	<b>Notations . . . . .</b>	<b>125</b>
	<b>Index . . . . .</b>	<b>129</b>

## Introduction

Finite automata were introduced in the 1940s as a mathematical model for nervous systems in living creatures. In course of time they were considered more and more as a fundamental model of computation. This led to the so-called theory of finite automata and regular languages which is one of the oldest theories in computer science.

In general, computer science investigates the computational possibilities of machines. In order to come close to reality one chooses machines that consist of a control unit and a memory. While the theory of Turing machines considers very powerful objects having an infinite memory, the theory of finite automata deals with very simple and restricted machines without memory. So in this sense we can say that (the power of) a real computer is something between a finite automaton and a Turing machine.

The concept of a finite automaton is a very simple one, and it can be explained even to a nonspecialist: We can think of a finite automaton as a machine that consists of some light bulbs (more precisely, one blue, some red and some green light bulbs) and a keypad. This machine behaves in such a way that at any point in time there is exactly one bulb that is on. When we switch on this machine then the blue bulb is on. When we press a key on the keypad then the automaton either does nothing or it switches to some other bulb (i.e., the current bulb is switched off and at the same time a new bulb is switched on). This behavior is exactly determined by the key that was pressed and by the bulb that was on. So whenever bulb  $X$  is on and key  $Y$  is pressed then always the same happens. Hence, when we switch on this machine and when we press a sequence of keys then we end in a situation where exactly one bulb is on. If this bulb is green then we say that this sequence of keys (i.e., this sequence of letters) is accepted, otherwise we say that it is rejected.

From the theoretical point of view finite automata can be characterized in many different ways. Among other things, the following characterizations are known: Kleene's theorem [Kle56] states that a language can be described by a regular expression if and only if it is recognizable by a finite automaton (which in turn is equivalent to saying that this language is accepted by a finite monoid). Therefore, languages accepted by a finite automaton are called *regular languages*. Büchi and Trakhtenbrot [Büc60, Tra61] showed that these are exactly the languages that can be described by a sentence of a certain monadic second-order logic.

In this thesis we consider permutationfree automata—a certain type of finite automata. Correspondingly, they have stronger characterizations than finite automata: McNaughton and Papert [MP71] showed that a minimal finite automaton is permutationfree if and only if its syntactic monoid is aperiodic. Schützenberger's theorem [Sch65] states that this is equivalent to saying that the accepted language can be described by a starfree

regular expression. Here, in contrast to general regular expressions, the iteration (Kleene star) is not allowed, i.e., these expressions consist only of letters, Boolean operations and concatenations. The subclass of regular languages that can be described by starfree regular expressions is called the class of *starfree languages* (for an overview we refer to [Pin95, Pin96b, Tho96]). This is exactly the class of languages that are definable by a sentence of a certain first-order logic [MP71].

Hence, by [Sch65, MP71] a regular language is starfree if and only if its minimal finite automaton is permutationfree. So for a given finite automaton we can compute the minimal equivalent automaton and we can test whether it is permutationfree. Since all these steps are effective this shows that one can decide whether a given finite automaton accepts a starfree language. In other words, the class of starfree languages is decidable.

We already mentioned that starfree languages are built up from letters by the use of Boolean operation on the one hand and concatenation on the other hand. Here the Boolean operations represent the *combinatorial* aspect and concatenations are responsible for the *sequential* aspect of the language. So in general, a starfree language is defined by the alternating use of both aspects. Brzozowski and Cohen [CB71] had the idea to count the minimal number of alternations that are inevitable to define a certain starfree language. This natural complexity measure of starfree languages is called the *dot-depth*. The corresponding question whether there exists an algorithm that determines the dot-depth of a given starfree language is known as the *dot-depth problem*.

Although finite automata—and in particular permutationfree automata—has this simple structure and although the membership problem for starfree languages is decidable, the dot-depth problem is still open. It is considered as one of the most famous and most difficult problems in the theory of finite automata [Pin98].

If we combine all languages with dot-depth  $n$  to a class  $\mathcal{B}_n$  (which is also called the  $n$ -th level) then this leads to the so-called dot-depth hierarchy [CB71]. In addition to this hierarchy we consider also the closely related Straubing-Thérien hierarchy [Str81, Thé81, Str85]. Since both hierarchies emerge when counting nested concatenations in starfree regular expressions they are also called *concatenation hierarchies*.

We state one of the many possibilities to define these concatenation hierarchies. Let  $A$  be some finite alphabet with  $|A| \geq 2$  (the hierarchies collapse in the unary case). The set of all words (respectively, nonempty words) over  $A$  is denoted by  $A^*$  (respectively,  $A^+$ ). For a class  $\mathcal{C}$  of languages let  $\text{Pol}(\mathcal{C})$  be its polynomial closure, i.e., the closure under finite union and concatenation, and denote by  $\text{BC}(\mathcal{C})$  its Boolean closure (taking complements w.r.t.  $A^+$  since we consider languages from  $A^+$ ). The classes  $\mathcal{L}_{n/2}$  of the Straubing-Thérien hierarchy and the classes  $\mathcal{B}_{n/2}$  of the dot-depth hierarchy are defined as follows.

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

By definition, all these classes are closed under union and it is known that they are also closed under intersection and under taking residuals [Arf91, PW97]. Both hierarchies are



strict [BK78, Tho84], closely related to each other [BK78, Str85, PW97, Sch01, PW01], and both exhaust the class of starfree languages [Eil76]. They formalize the dot-depth problem in terms of their hierarchy classes, i.e., the minimal level containing a given language.

If we consider a fixed class in these hierarchies then again the question for the decidability of the corresponding membership problems arises. Up to now, only the levels  $1/2$ ,  $1$  and  $3/2$  of both hierarchies are known to be decidable [Sim75, Kna83, Arf91, PW97, GS00a], while the question is open for any other level. Partial results are known for level  $2$  and level  $5/2$  of the Straubing-Thérien hierarchy—both levels are decidable if a two-letter alphabet is considered [Str88, GS00b].

So at the moment one cannot answer the decidability of a level  $n/2$  for  $n \geq 4$ . Nevertheless, the answer for a dot-depth class is the same as for a Straubing-Thérien class. More precisely, Straubing [Str85] showed that level  $n$  of the dot-depth hierarchy is decidable if and only if level  $n$  of the Straubing-Thérien hierarchy is decidable (for integers  $n$ ). Recently, Pin and Weil [PW01] proved that this is also true for the levels  $n + 1/2$ .

Corresponding to the various characterizations of regular languages and starfree languages there are also characterizations for the single levels of the dot-depth hierarchy and of the Straubing-Thérien hierarchy. The dot-depth hierarchy is related to the first-order logic  $\text{FO}[\langle, \min, \max, S, P]$ , and the Straubing-Thérien hierarchy corresponds to the first-order logic  $\text{FO}[\langle]$  in the following sense. Both logics have the unary relations for the alphabet symbols from  $A$  and the binary relation  $\langle$ . Moreover,  $S$  (respectively,  $P$ ) is the successor (respectively, predecessor) function and  $\min$ ,  $\max$  are constants. For a fixed first-order logic let  $\Sigma_n$  be the class of languages that can be described by a sentence with at most  $n - 1$  quantifier alternations, starting with an existential quantifier. It has been proved by Thomas [Tho82], and Perrin and Pin [PP86] that  $\Sigma_n$  formulas of  $\text{FO}[\langle, \min, \max, S, P]$  (respectively,  $\text{FO}[\langle]$ ) describe just the languages from  $\mathcal{B}_{n-1/2}$  (respectively,  $\mathcal{L}_{n-1/2}$ ).

Once again let us return to the characterization of starfree languages by permutationfree automata [Sch65, MP71]. We say that a finite automaton has a nontrivial permutation if and only if there exist a nonempty word  $w$ , an integer  $l \geq 2$  and distinct states  $r_1, r_2, \dots, r_l$  such that on input  $w$  the automaton moves from  $r_l$  to  $r_1$  and from  $r_i$  to  $r_{i+1}$  for  $1 \leq i \leq l-1$ . A minimal automaton is called permutationfree if and only if it does not have a nontrivial permutation. So [Sch65, MP71] shows a characterization of starfree languages in terms of structural properties in the transition graphs of automata. More precisely, a language accepted by a finite automaton  $\mathcal{M}$  is starfree if and only if  $\mathcal{M}$  does *not* have a nontrivial permutation. Hence the characterization is such that a certain pattern is forbidden in the automata, and therefore this is called a *forbidden-pattern characterization* of starfree languages.

Usually, forbidden-pattern characterizations imply the decidability of the characterized class, since we only have to test whether the forbidden-pattern occurs in an automaton. So a forbidden-pattern characterization says more than just decidability. It relates the *absence* of a certain pattern in the automaton with the *existence* of an expression describing the accepted language. This means that such a characterization shows us the structure in the automaton that cannot be expressed by the characterized class (and therefore what causes a language to be not in this class).

Interestingly, most of the decidability results for classes of concatenation hierarchies go along with forbidden-pattern characterizations. In this thesis we follow this idea and introduce two hierarchies that consist of classes of starfree languages, so-called *forbidden-pattern classes*. These are decidable classes since they are defined via forbidden-patterns. Then we show that these decidable forbidden-pattern classes contain the classes of the dot-depth hierarchy and the classes of the Straubing-Thérien hierarchy. Using the *technique of word extensions* we prove that the classes  $\mathcal{B}_{1/2}$ ,  $\mathcal{B}_{3/2}$ ,  $\mathcal{L}_{1/2}$  and  $\mathcal{L}_{3/2}$  even coincide with the respective forbidden-pattern classes. This implies their decidability. Moreover, with the same technique we also show that the Boolean hierarchies over  $\mathcal{L}_{1/2}$  and over  $\mathcal{B}_{1/2}$  are decidable.

At this point we want to make a general bibliographic remark. The sections 1.1 and 1.2 introduce basic definitions and concepts which are known from the literature. The remaining parts of chapter 1 (i.e., the sections 1.3 and 1.4) consist of work done by the author. Chapter 2 summarizes known results for concatenation hierarchies. The theory of forbidden-patterns in chapter 3 was developed by the author in joint work with Heinz Schmitz, Würzburg, and it is also part of his thesis [Sch01].

The main results in chapter 4 go back to the following authors: The decidability of  $\mathcal{B}_{1/2}$ ,  $\mathcal{L}_{1/2}$  and  $\mathcal{L}_{3/2}$  were first shown in [Arf91, PW97]. For the Boolean hierarchy over  $\mathcal{L}_{1/2}$  this is known from [SW98], and for the Boolean hierarchy over  $\mathcal{B}_{1/2}$  this is due to the author [Gla99]. The decidability of  $\mathcal{B}_{3/2}$  was shown by the author in joint work with Heinz Schmitz [GS00a] and it is also part of his thesis [Sch01]. However, in chapter 4 we use a new approach which differs from [Arf91, PW97, GS00a, Sch01]. More precisely, we develop a technique of word extensions which makes it possible to obtain all 6 decidability results in a uniform way.

## Chapter One

At the beginning of this introductory chapter we give some basic definitions and notations. Then we define the fundamental notion of well partial ordered sets and prove some of their properties.

Another part of this chapter provides a combinatorial tool that allows to partition words of arbitrary length into factors of bounded length such that every second factor  $u$  leads to a loop with label  $u$  in a given finite automaton. So from an algebraic point of view these words  $u$  are idempotents with respect to the given finite automaton.

Finally, we introduce the word extensions  $\langle_v^{0,k}$  and  $\langle_v^{1,k}$ . We prove some basic properties and investigate the  $\langle_v^{1,k}$  upward closure of certain languages. These and other word extensions will play a central role in the forthcoming proofs for decidability results.

## Chapter Two

Here we give definitions for the dot-depth hierarchy and for the closely related Straubing-Thérien hierarchy. We formulate the dot-depth problem, prove some easy inclusion relations and state that both hierarchies exhaust the class of starfree languages.

We also discuss the use of alternative definitions for the considered concatenation hierarchies. It will turn out that this makes no difference when looking at the essential decidability questions of these hierarchies.

Finally, we mention known decidability results and known forbidden-pattern characterizations. We compare the known forbidden-patterns of the lower levels to get an idea of how a general structure for concatenation hierarchies could look like.

## Chapter Three

This chapter is devoted to forbidden-pattern characterizations which are results of the following type: “A language  $L$  belongs to a class  $\mathcal{C}$  if and only if the accepting finite automaton does not have subgraph  $\mathbb{P}$  in its transition graph”.

If we compare the known forbidden-pattern characterizations for  $\mathcal{L}_{1/2}$  and  $\mathcal{L}_{3/2}$  we observe that the patterns for  $\mathcal{L}_{1/2}$  act as building blocks in the patterns for  $\mathcal{L}_{3/2}$ . We find this observation confirmed, if we compare the patterns for  $\mathcal{B}_{1/2}$  with the characterization of  $\mathcal{B}_{3/2}$ . This motivates the introduction of an iteration rule IT on patterns, which continues the observed formation procedure.

Starting from an initial class of patterns our iteration rule IT generates classes of more complicated patterns. If we forbid these classes of patterns in finite automata then this defines classes of language—the so-called forbidden-pattern classes. The main technical result of this chapter relates in a general way the iteration rule IT to the polynomial closure operation Pol.

We apply our results to particular initial classes of patterns which correspond to the first levels of the dot-depth hierarchy and the Straubing-Thérien hierarchy, respectively. We obtain strict and decidable hierarchies of forbidden-pattern classes  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{B}})$  and  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$  which exhaust the class of starfree languages. Then we prove that these classes of languages contain the corresponding classes of the concatenation hierarchies.

Finally, we provide more structural similarities between the classes of the concatenation hierarchies and the forbidden-pattern classes: All hierarchies show the same inclusion structure. Moreover, typical languages that separate the classes of the concatenation hierarchies also separate levelwise our forbidden-pattern classes.

## Chapter Four

In this chapter we restrict ourselves to the levels  $n + 1/2$  of the dot-depth hierarchy and of the Straubing-Thérien hierarchy. We prove the decidability of the levels 1/2 and 3/2 of both hierarchies in terms of forbidden-pattern characterizations for these classes. Furthermore, we show the decidability of the Boolean hierarchies over  $\mathcal{B}_{1/2}$  and  $\mathcal{L}_{1/2}$ .

More precisely, from chapter 3 we know the inclusions  $\mathcal{B}_{n+1/2} \subseteq \mathcal{FP}(\mathbb{P}_n^{\mathcal{B}})$  and  $\mathcal{L}_{n+1/2} \subseteq \mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$ . For the reverse inclusions (i.e., the more difficult ones in our forbidden-pattern characterizations) we use a technique which is based on word extensions. With this technique it is possible to treat the classes  $\mathcal{L}_{1/2}$ ,  $\mathcal{B}_{1/2}$ ,  $\mathcal{L}_{3/2}$  and  $\mathcal{B}_{3/2}$  in a uniform way. Moreover,

we can use these word extensions to prove the decidability of the above mentioned Boolean hierarchies.

## Summary

Forbidden-patterns combined with word extensions turn out to be a useful tool to attack the decidability of the dot-depth classes. With it we are able to prove the decidability of  $\mathcal{B}_{3/2}$  and the decidability of the Boolean hierarchy over  $\mathcal{B}_{1/2}$ —two results which were not known before.

The disadvantages of this approach are its cumbersome notations and proofs. It seems to be likely that the forbidden-patterns for higher levels of the dot-depth hierarchy (if they exist at all) become more and more complex. So we have to expect that the forbidden-pattern approach, applied to higher levels, will lead to even more cumbersome proofs. Unfortunately, the same holds for the algebraic automata theory since patterns in automata and equations/inequalities in semigroups are basically the same things. However, up to now the algebraic approach and the forbidden-pattern approach are the most successful approaches to decidability results for the dot-depth classes.

Recently, a new, purely logical approach to decidability questions for concatenation hierarchies was proposed in [Sel01]. With this approach it is possible to give short proofs for the decidability of  $\mathcal{L}_{1/2}$ ,  $\mathcal{L}_1$ ,  $\mathcal{B}_1$ , and the classes of the Boolean hierarchy over  $\mathcal{L}_{1/2}$ .

The theory of word extensions for  $\mathcal{B}_{1/2}$  in section 4.1 led to the decidability of the Boolean hierarchy over this class (see section 4.2). In section 4.3 we develop a very similar theory, this time for  $\mathcal{B}_{3/2}$ . This could be a promising starting point to attack the decidability of the Boolean hierarchy over  $\mathcal{B}_{3/2}$ . Note that this would be a remarkable step towards the decidability of  $\mathcal{B}_2$ .

## Publications

Parts of this thesis are published in the following papers.

- [GS00a] C. Glaßer and H. Schmitz, *Languages of dot-depth 3/2*. Proceedings of the 17th Symposium on Theoretical Aspects of Computer Science, 2000.
- [GS00c] C. Glaßer and H. Schmitz, *The Boolean structure of dot-depth one*. To appear in Journal of Automata, Languages and Combinatorics, 2001.
- [GS00d] C. Glaßer and H. Schmitz, *Decidable hierarchies of starfree languages*. Proceedings of the 20th Conference on the Foundations of Software Technology and Theoretical Computer Science, 2000.
- [GS01] C. Glaßer and H. Schmitz, *Level 5/2 of the Straubing-Thérien hierarchy for two-letter alphabets*. Preproceedings of the 5th Conference on Developments in Language Theory, 2001.

Moreover, the technical reports [Gla98, Gla99, GS99, GS00b] contain parts of this thesis.

# 1. Preliminaries

This chapter has an introductory character. We start in section 1.1 with some basic definitions and notations. Then in section 1.2 we deal with the fundamental notion of well partial ordered sets.

Section 1.3 provides a combinatorial tool that allows to partition words of arbitrary length into factors of bounded length such that every second factor  $u$  is an idempotent, i.e.,  $u$  leads to a loop with label  $u$  in a given finite automaton.

Finally, in section 1.4 we introduce the word extensions  $\prec_v^{0,k}$  and  $\prec_v^{1,k}$ . Both can be also considered as binary relations on the set of all words. These and other word extensions will play a central role in the forthcoming proofs for decidability results. The major part of this section investigates the  $\prec_v^{1,k}$  upward closure of certain languages.

## 1.1 Definitions and Notations

Throughout this thesis we fix some arbitrary finite alphabet  $A$  with  $|A| \geq 2$ . The empty word is denoted by  $\varepsilon$ , the set of all words over  $A$  (including the empty word) is denoted by  $A^*$  and the set of all nonempty words over  $A$  is denoted by  $A^+$ . The length of a word  $w$  is denoted by  $|w|$ . Moreover, for  $k \geq 0$  we use the following notations for sets of words.

$$\begin{aligned} A^k &=_{\text{def}} \{ w \in A^* \mid |w| = k \} \\ A^{*\leq k} &=_{\text{def}} \{ w \in A^* \mid |w| \leq k \} \\ A^{*\geq k} &=_{\text{def}} \{ w \in A^* \mid |w| \geq k \} \\ A^{+\leq k} &=_{\text{def}} \{ w \in A^+ \mid |w| \leq k \} \\ A^{+\geq k} &=_{\text{def}} \{ w \in A^+ \mid |w| \geq k \} \end{aligned}$$

Languages are considered as subsets of  $A^+$ , and therefore complementation is taken with respect to  $A^+$ . So for a class  $\mathcal{C}$  of languages of  $A^+$ ,  $\text{co}\mathcal{C} =_{\text{def}} \{ A^+ \setminus L \mid L \in \mathcal{C} \}$  denotes the set of complements w.r.t.  $A^+$ .

Let  $k \geq 0$  and  $w \in A^*$  with  $w = a_1 \cdots a_n$  for alphabet letters  $a_i \in A$ . If  $v = a_i a_{i+1} \cdots a_{j-1}$  for  $1 \leq i \leq j \leq n+1$  then we call  $v$  a factor of  $w$ . If  $v = a_{i_1} a_{i_2} \cdots a_{i_m}$  for  $m \geq 0$  and  $1 \leq i_1 < i_2 < \cdots < i_m \leq n$  then  $v$  is called a subword of  $w$  (in this case we write  $v \preceq w$  for short). Define  $\alpha_k(w)$  to be the set of factors of length  $k+1$  that occur in  $w$ , i.e.,  $\alpha_k(w) =_{\text{def}} \{ v \in A^{k+1} \mid v \text{ is a factor of } w \}$ . For a language  $L \subseteq A^*$ , let  $w^{-1}L =_{\text{def}} \{ v \in A^* \mid vw \in L \}$  be the left residual of  $L$ , and let

$Lw^{-1} =_{\text{def}} \{v \in A^* \mid vw \in L\}$  be the right residual of  $L$ . With  $A^{-j}w$  (respectively,  $wA^{-j}$ ) we denote the word that emerges from  $w$  when deleting the first  $j$  (respectively, last  $j$ ) letters of  $w$ . If  $j \geq |w|$  then we set  $A^{-j}w = wA^{-j} = \varepsilon$ . The  $k$ -prefix and the  $k$ -suffix of  $w$  are defined as follows.

$$\begin{aligned} \mathfrak{p}_k(w) &=_{\text{def}} \begin{cases} a_1a_2 \cdots a_k & : \text{ if } k \leq n \\ \varepsilon & : \text{ otherwise} \end{cases} \\ \mathfrak{s}_k(w) &=_{\text{def}} \begin{cases} a_{n-k+1}a_{n-k+2} \cdots a_n & : \text{ if } k \leq n \\ \varepsilon & : \text{ otherwise} \end{cases} \end{aligned}$$

Moreover, for  $1 \leq i \leq j \leq n+1$  let  $w[i, j] =_{\text{def}} a_i a_{i+1} \cdots a_{j-1}$ .

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. Since we look at languages of  $A^+$  we take complements with respect to  $A^+$ .

A deterministic finite automaton (DFA)  $\mathcal{M}$  is given by  $\mathcal{M} = (A, S, \delta, s_0, S')$ , where  $A$  is its input alphabet,  $S$  is its set of states,  $\delta : A \times S \rightarrow S$  is its total transition function,  $s_0 \in S$  is the initial state and  $S' \subseteq S \setminus \{s_0\}$  is the set of accepting states. We denote by  $L(\mathcal{M})$  the language accepted by  $\mathcal{M}$  (note that  $L(\mathcal{M}) \subseteq A^+$  since we demand  $s_0 \notin S'$ ). As usual, we extend transition functions to input words, and we denote by  $|\mathcal{M}|$  the number of states of  $\mathcal{M}$ . For a word  $z \in A^*$  we use  $s_1 \xrightarrow{z} s_2$  as an abbreviation for  $\delta(s_1, z) = s_2$ . With  $s_1 \xrightarrow{z} s_2$  we mean that there exists some  $z \in A^*$  such that  $s_1 \xrightarrow{z} s_2$ . Moreover, we write  $s_1 \xrightarrow{z} +$  (respectively,  $s_1 \xrightarrow{z} -$ ) if there exists a state  $s_2 \in S'$  (respectively,  $s_2 \in S \setminus S'$ ) with  $s_1 \xrightarrow{z} s_2$ . A minimal DFA  $\mathcal{M}$  is a DFA such that for all  $\mathcal{M}'$  with  $L(\mathcal{M}) = L(\mathcal{M}')$  it holds that  $|\mathcal{M}| \leq |\mathcal{M}'|$ .

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  $\delta^w : S \rightarrow S$  with  $\delta^w(s) =_{\text{def}} \delta(s, w)$ . Say that a total mapping  $\delta' : S \rightarrow S$  leads to a  $v$ -loop (respectively, leads to some structure) if and only if  $\delta'(s)$  has a  $v$ -loop (respectively, has this structure) for all  $s \in S$ . We may also say for short that a word  $w \in A^*$  leads to a  $v$ -loop (respectively, leads to some structure) if  $\delta^w$  does so. Moreover, for  $v, w \in A^*$  we write  $v \sim_{\mathcal{M}} w$  if and only if  $\delta^v = \delta^w$ . A word  $u \in A^*$  is called an idempotent for  $\mathcal{M}$  if and only if  $\delta^u = \delta^{uu}$ .

As usual let P (respectively, NP) be the class of languages that can be accepted by a Turing machine in deterministic polynomial time (respectively, nondeterministic polynomial time). Moreover, the class of languages that can be accepted in nondeterministic logarithmic space (respectively, deterministic logarithmic space, deterministic polynomial space) is denoted by NL (respectively, L, PSPACE). For more information about these complexity classes see, e.g., [Pap94].

An obvious property of DFAs is that they run into loops after a small number of successive  $w$ 's in the input.

**Proposition 1.1.** *Let  $w \in A^*$  and  $r \geq 1$ , then  $w^r$  leads to a  $w^{r!}$ -loop in every DFA  $\mathcal{M}$  with  $|\mathcal{M}| \leq r$ .*

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

We recall the following fundamental theorem concerning starfree languages.

**Theorem 1.2** ([Sch65, MP71]). *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a minimal DFA. Then  $L(\mathcal{M})$  is not starfree if and only if there exist a word  $w \in A^+$ , some  $l \geq 2$  and distinct states  $r_1, r_2, \dots, r_l \in S$  such that*

$$r_1 \xrightarrow{w} r_2 \xrightarrow{w} \dots \xrightarrow{w} r_l \xrightarrow{w} r_1.$$

We call a minimal DFA  $\mathcal{M}$  permutationfree if it has the above property. Deciding this property for a given  $\mathcal{M}$  is known to be PSPACE-complete [CH91] (in Remark 3.31 we make precise how we think of a DFA as an input to a Turing machine).

Let  $w \in A^*$  with  $w = a_1 \dots a_n$  for alphabet letters  $a_i \in A$ . If  $w = x v z$  for words  $x, v, z \in A^*$  then we call  $x v z$  a decomposition of  $w$ . If we speak about the “factor  $v$  of the decomposition  $w = x v z$ ” then of course we do not mean an arbitrary appearance of  $v$  in  $w$  but exactly that appearance starting at the  $(|x| + 1)$ -st letter and ending at the  $(|x| + |v|)$ -th letter of  $w$ . Let  $w = x' v' z'$  be another decomposition. Then the formulations given in the table below have the following meaning.

Formulation	Meaning
“the factor $v$ of the decomposition $w = x v z$ is contained in the factor $v'$ of the decomposition $w = x' v' z'$ ”	$ x  \geq  x' $ and $ z  \geq  z' $
“the factor $v$ of the decomposition $w = x v z$ overlaps the factor $v'$ of the decomposition $w = x' v' z'$ ”	$ x v  >  x' $ and $ x' v'  >  x $

We write  $\mathcal{P}(B)$  for the power set of an arbitrary set  $B$ . When we speak about finite unions of sets then this includes the empty union (which yields the empty set). For a rational number  $r$  we define  $\lfloor r \rfloor$  as the greatest integer that is less than or equal to  $r$ .

**Definition 1.3** ([KSW87, CGH<sup>+</sup>88]). *Let  $\mathcal{C}$  be a class of languages closed under union and intersection. The Boolean hierarchy over  $\mathcal{C}$  is the family of classes  $\mathcal{C}(l)$  and  $\text{co}\mathcal{C}(l)$  for  $l \geq 1$ , where  $\mathcal{C}(l)$  is the class of languages  $L$  that can be written as  $L = L_1 \setminus (L_2 \setminus (\dots \setminus L_l))$  for some  $L_1, L_2, \dots, L_l \in \mathcal{C}$  with  $L_1 \supseteq L_2 \supseteq \dots \supseteq L_l$ .*

**Lemma 1.4** ([KSW87, CGH<sup>+</sup>88]). *Let  $\mathcal{C}$  be closed under union and intersection. Then  $\mathcal{C}(l) \cup \text{co}\mathcal{C}(l) \subseteq \mathcal{C}(l+1) \cap \text{co}\mathcal{C}(l+1)$  for  $l \geq 1$  and the Boolean closure  $\text{BC}(\mathcal{C}) = \bigcup_{l \geq 1} \mathcal{C}(l)$ .*

It is known from these papers that every class defined via a fixed but arbitrary Boolean combination of the languages from  $\mathcal{C}$  coincides with one of the classes  $\mathcal{C}(l)$  or  $\text{co}\mathcal{C}(l)$ .

## 1.2 Well Partial Ordered Sets

In order to prove decidability results for concatenation hierarchies we will introduce word extensions (see chapter 4). They can be also considered as binary relations on the set of words  $A^*$ . Since these relations are reflexive, transitive and antisymmetric they make the set of words to a partial ordered set. We will show that some of these relations make  $A^*$  even to a well partial ordered set. This means that for every nonempty subset of  $A^*$  the set of minimal elements in this subset is nonempty and finite. Below we give formal definitions for co-ideals and well partial ordered sets. We show that every co-ideal of a well partial ordered set is finitely generated (this fact is important for the proofs of the decidability results in chapter 4).

**Definition 1.5.** *Let  $S$  be a set and let  $<$  be a binary relation on  $S$ . A subset  $I \subseteq S$  is called a  $<$  co-ideal if and only if for all  $s \in S$  and  $x \in I$  with  $x < s$  it holds that  $s \in I$ .*

**Definition 1.6.** *Let  $S$  be a set and let  $<$  be a binary relation on  $S$ . For  $s \in S$  and  $T \subseteq S$  we define the  $<$  upward closure of  $s$  and  $T$  as*

$$\begin{aligned} \langle s \rangle_{<} &=_{\text{def}} \{s\} \cup \{t \in S \mid (\exists n \geq 0, t_0, \dots, t_n \in S)[s < t_0 < \dots < t_n = t]\} \quad \text{and} \\ \langle T \rangle_{<} &=_{\text{def}} \bigcup_{t \in T} \langle s \rangle_{<}. \end{aligned}$$

We refer to the  $<$  upward closure of  $s$  and  $T$  also as the  $<$  co-ideal generated by  $s$  and  $T$ . Correspondingly, we say that a  $<$  co-ideal  $I \subseteq S$  is finitely generated if and only if  $I = \langle D \rangle_{<}$  for a finite set  $D \subseteq S$ . Note that if  $<$  is reflexive and transitive then we can simplify the definition above and we obtain  $\langle s \rangle_{<} = \{t \in S \mid s < t\}$ .

**Definition 1.7.** *Let  $(S, \leq)$  be an ordered set (i.e.,  $\leq$  is a binary, reflexive and transitive relation on  $S$ ). We call  $(S, \leq)$  a well partial ordered set (wpos, for short) if and only if all  $T \subseteq S$  satisfy the following conditions.*

1.  $T$  does not have an infinite, strictly descending chain (i.e., elements  $t_0, t_1, \dots \in T$  such that  $t_{i+1} \leq t_i$  and  $t_{i+1} \neq t_i$  for  $i \geq 0$ ).
2.  $T$  does not have an infinite set of pairwise incomparable words (i.e., an infinite  $D \subseteq T$  such that  $t_1 \not\leq t_2$  for all  $t_1, t_2 \in D$  with  $t_1 \neq t_2$ ).

This is equivalent to saying that for every nonempty subset of  $S$  the set of minimal elements with respect to  $\leq$  is nonempty and finite [CK96]. For several equivalent properties, which may be used for the definition of well partial ordered sets, see [SS83, CK96].

**Proposition 1.8.** *Let  $(S, \leq)$  be an ordered set and  $T \subseteq S$ . Then  $\langle T \rangle_{\leq}$  is finitely generated if it satisfies the following conditions.*

1.  $T$  does not have an infinite, strictly descending chain.
2.  $T$  does not have an infinite set of pairwise incomparable words.



*Proof.* Let  $M$  be the set of minimal elements in  $T$ . Since different minimal elements of  $T$  are incomparable it holds that  $M$  is finite.

Suppose  $\langle M \rangle_{\leq} \neq \langle T \rangle_{\leq}$ . Since  $\langle M \rangle_{\leq} \subseteq \langle T \rangle_{\leq}$  there exists an  $s \in \langle T \rangle_{\leq} \setminus \langle M \rangle_{\leq}$ . Hence there exists some  $t_0 \in T \setminus \langle M \rangle_{\leq}$  with  $t_0 \leq s$  (note that  $\leq$  is reflexive and transitive). It follows that  $t_0$  is not a minimal element in  $T$ . This implies that there exists a  $t_1 \in T \setminus \langle M \rangle_{\leq}$  with  $t_1 \neq t_0$  and  $t_1 \leq t_0$ . Analogously we obtain elements  $t_2, t_3, \dots \in T \setminus \langle M \rangle_{\leq}$  such that  $t_{i+1} \neq t_i$  and  $t_{i+1} \leq t_i$  for  $i \geq 0$ . So we have found an infinite, strictly descending chain. This contradicts our assumption and it follows  $\langle M \rangle_{\leq} = \langle T \rangle_{\leq}$ .  $\square$

**Proposition 1.9.** *If  $(S, \leq)$  is a wpos and  $T \subseteq S$  then  $\langle T \rangle_{\leq}$  is finitely generated.*

*Proof.* This follows from Definition 1.7 and Proposition 1.8.  $\square$

### 1.3 The Loop-Lemma

Let  $\mathcal{M}$  be a DFA and consider an arbitrary decomposition  $w = v_0 v_1 \cdots v_n$  of a nonempty word  $w$  (we demand also that the words  $v_i$  are nonempty). A useful tool in further proofs is the fact that if  $n$  is large enough then we can find a factor  $u =_{\text{def}} v_i v_{i+1} \cdots v_j$  that leads to  $u$ -loops in  $\mathcal{M}$ . This can be compared to the algebraic notion of idempotents in finite semigroup theory. In particular, if we state the result for factors of length one (i.e., all  $v_i \in A$ ), this means that in every sufficiently ‘long’ word  $w$  we find a ‘short’ nonempty factor  $u$  such that  $u$  is an idempotent for  $\mathcal{M}$  (i.e.,  $\delta^{uu} = \delta^u$  where  $\delta$  is the transition function of  $\mathcal{M}$ ).

It is important here to analyze the number of blocks needed to find such a factor (i.e.,  $n + 1$  in our example). We will prove that this number can be bounded by a function in  $|\mathcal{M}|$ . For this end, we define the following function.

**Definition 1.10.** *For a DFA  $\mathcal{M}$  let  $\mathcal{I}_{\mathcal{M}} =_{\text{def}} (|\mathcal{M}| + 1)^{(|\mathcal{M}|+1)^{(|\mathcal{M}|+1)}}$ .*

We first show with a rather rough estimation, that  $\mathcal{I}_{\mathcal{M}}$  does not become too small if we repeatedly divide it by  $|\mathcal{M}|^{|\mathcal{M}|}$ . This will make the proof of Lemma 1.13 below better readable.

**Proposition 1.11.** *Let  $\mathcal{M}$  be a DFA,  $n =_{\text{def}} |\mathcal{M}|$ ,  $m_1 =_{\text{def}} \lfloor \mathcal{I}_{\mathcal{M}}/2 \rfloor$  and  $m_{i+1} =_{\text{def}} \lfloor m_i/n^n \rfloor - 1$  for  $i \geq 1$ . For  $1 \leq i \leq n^n + 1$  it holds that*

$$m_i \geq (2n^n)^{(n^n+3-i)}.$$

*Proof.* We will prove the lemma by induction on  $i$  with  $1 \leq i \leq n^n + 1$ . For the induction base let  $i = 1$ . We distinguish two cases, first suppose  $n = 1$ . By definition of  $\mathcal{I}_{\mathcal{M}}$ , we have  $m_1 = 8 = (2n^n)^{(n^n+3-1)}$ . Now let  $n \geq 2$ . By the binomial theorem we have in this case  $(n+1)^{n+1} \geq n^{n+1} + (n+1)n^n + (n+1)n + 1 \geq n^{n+1} + 2n + 2$  and  $n^n + n \cdot n^{n-1} \leq (n+1)^n$ . So the following estimation shows the induction base.

$$\begin{aligned}
(2n^n)^{(n^n+3-1)} &= (n^n + n \cdot n^{n-1})^{(n^n+3-1)} &\leq (n+1)^{n(n^n+2)} \\
& &\leq (n+1)^{((n+1)^{(n+1)}-2)} \\
& &\leq (n+1)^{((n+1)^{(n+1)}-1)} - 1 \\
& &\leq \frac{(n+1)^{((n+1)^{(n+1)})}}{2} - 1 \\
& &\leq \lfloor \mathcal{I}_{\mathcal{M}}/2 \rfloor = m_1
\end{aligned}$$

For the induction step, suppose that we have already shown  $m_k \geq (2n^n)^{(n^n+3-k)}$  with  $1 \leq k < n^n + 1$ . By definition,  $m_{k+1} = \lfloor m_k/n^n \rfloor - 1$ . From the induction hypothesis we obtain

$$m_{k+1} \geq \frac{(2n^n)^{(n^n+3-k)}}{n^n} - 2.$$

Since  $n \geq 1$  and  $k < n^n + 1$  we have  $(2n^n)^{(n^n+3-(k+1))}/n^n \geq 4$ . It follows that

$$m_{k+1} \geq \frac{(2n^n)^{(n^n+3-k)}}{n^n} - 2 = 2 \cdot (2n^n)^{(n^n+3-(k+1))} - 2 \geq (2n^n)^{(n^n+3-(k+1))}.$$

□

The key argument for Lemma 1.13 below is the iterated use of the fact that there is only a finite number of mappings  $\delta' : S \rightarrow S$  when a finite set  $S$  is given. We isolate the iteration step in the following lemma. Let a word  $v$  be given with a decomposition  $v = v_0v_1 \cdots v_l$  for sufficiently large  $l$ . Among the mappings  $\delta^{v_1 \cdots v_j}$  some coincide if  $l$  is large enough. Suppose for instance, there are  $x, y, z, v'$  such that  $v = xyzv'$  and  $\delta^x = \delta^{xy} = \delta^{xyz}$ . Then  $\delta^x$  leads to a  $y$ -loop and also to a  $z$ -loop. We repeat this selection procedure on the now coarser decomposition  $xyzv' = v = v_1v_2 \cdots v_l$ , and collect the hereby encountered mappings in the set  $\Delta$ .

In order to make this precise, let  $v_0, v_1, \dots, v_l \in A^+$  and define  $v(i, j) =_{\text{def}} v_i v_{i+1} \cdots v_{j-1}$  for all  $0 \leq i < j \leq l + 1$ . We work with indices  $i_0, \dots, i_m$  in order to allow iterated applications.

**Lemma 1.12.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA and let  $v_0, v_1, \dots, v_l \in A^+$  be given. Furthermore, let  $0 \leq i_0 < i_1 < \cdots < i_m \leq l$  and suppose that  $\Delta$  is a set of total mappings  $\delta' : S \rightarrow S$  such that every  $\delta' \in \Delta$  leads to a  $v(i_j, i_{j+1})$ -loop for all  $0 \leq j < m$ . Then there exist indices  $i'_0 < i'_1 < \cdots < i'_n$  with  $n =_{\text{def}} \lfloor m/(\lceil |S|^{|S|} \rceil) \rfloor$  such that*

1.  $\{i'_0, i'_1, \dots, i'_n\} \subseteq \{i_0, i_1, \dots, i_m\}$ ,
2. if  $n \geq 1$  then every  $\delta' \in \Delta$  leads to a  $v(i'_0, i'_1)$ -loop, and
3. if  $n \geq 1$  then every  $\delta' \in \Delta \cup \left\{ \delta^{v(i'_0, i'_1)} \right\}$  leads to a  $v(i'_j, i'_{j+1})$ -loop for all  $1 \leq j < n$ .

*Proof.* First, set  $i'_0 =_{\text{def}} i_0$ . In particular this shows the lemma for  $n = 0$ . If  $n = 1$  we set  $i'_1 =_{\text{def}} i_1$  and we are done. Suppose  $n \geq 2$  and set  $\delta_{i_j} =_{\text{def}} \delta^{v(i_0, i_j)}$  for  $1 \leq j \leq m$ . Since there are at most  $|S|^{|S|}$  total mappings  $S \rightarrow S$ , there exist mappings that appear several times in the list  $\delta_{i_1}, \delta_{i_2}, \dots, \delta_{i_m}$ . From these mappings we choose a mapping  $\bar{\delta}$

that appears most frequently, say  $\bar{\delta}$  appears  $n'$  times. So  $n' \geq \lfloor m/(|S|^{|S|}) \rfloor = n$ . Let  $i'_1, i'_2, \dots, i'_n \in \{i_1, i_2, \dots, i_m\}$  such that  $i'_1 < i'_2 < \dots < i'_n$  and  $\bar{\delta} = \delta_{i'_j}$  for  $1 \leq j \leq n$ .

Since  $\{i'_1, i'_2, \dots, i'_n\} \subseteq \{i_1, i_2, \dots, i_m\}$  we see the first statement. By assumption, every  $\delta' \in \Delta$  leads to a  $v(i_j, i_{j+1})$ -loop for all  $0 \leq j < m$ . It follows that every  $\delta' \in \Delta$  leads also to a  $u$ -loop, where  $u$  is an arbitrary concatenation of words  $v(i_j, i_{j+1})$  with  $0 \leq j < m$ . Particularly, every  $\delta' \in \Delta$  leads to a  $v(i'_j, i'_{j+1})$ -loop for all  $0 \leq j < n$ . Thus the second statement follows, and we obtain also the third statement for  $\delta' \in \Delta$ .

It remains to show that  $\delta^{v(i'_0, i'_1)}$  leads to a  $v(i'_j, i'_{j+1})$ -loop for all  $1 \leq j < n$ . By the choice of  $\bar{\delta}$  we have that  $\delta^{v(i'_0, i'_j)} = \bar{\delta} = \delta^{v(i'_0, i'_{j+1})}$  for all  $1 \leq j < n$  and we see that  $\bar{\delta}$  leads to an  $v(i'_j, i'_{j+1})$ -loop for all  $1 \leq j < n$ . Since  $\delta^{v(i'_0, i'_1)} = \bar{\delta}$  the third statement follows.  $\square$

Note that the second statement in the previous lemma and also third statement for  $\delta' \in \Delta$  follow immediately from the first statement. We explicitly state them here to focus on what is important in the following proof. We use the same finiteness argument as before: the mapping we add to  $\Delta$  in Lemma 1.12 cannot always be a new mapping. So if the number of factors we start with is large enough to allow many applications of Lemma 1.12, then we find a mapping  $\delta^u$  that has already been added to  $\delta$  before, say  $\delta'$ . But this means by the second statement of Lemma 1.12 that  $\delta'$  leads to a  $u$ -loop, and hence  $\delta' = \delta^u = \delta^{uu}$ .

**Lemma 1.13.** *For every DFA  $\mathcal{M} = (A, S, \delta, s_0, S')$  and for all  $v_0, v_1, \dots, v_l \in A^+$  with  $l =_{\text{def}} \lfloor \mathcal{I}_{\mathcal{M}}/2 \rfloor$  there exist  $0 \leq g < h \leq l$  such that  $\delta^{uu} = \delta^u$  with  $u =_{\text{def}} v_g v_{g+1} \dots v_{h-1}$ .*

*Proof.* Let  $n =_{\text{def}} |\mathcal{M}|$ . Initially, let  $m^{(1)} =_{\text{def}} l$ ,  $\Delta^{(1)} =_{\text{def}} \emptyset$  and  $i_r^{(1)} =_{\text{def}} r$  for  $0 \leq r \leq l$ . We apply Lemma 1.12 the first time and obtain for  $n^{(1)} =_{\text{def}} \lfloor m^{(1)}/n^n \rfloor$  indices  $i'_r{}^{(1)}$  with  $0 \leq r \leq n^{(1)}$  such that

1.  $\left\{ i'_r{}^{(1)} \mid 0 \leq r \leq n^{(1)} \right\} \subseteq \left\{ i_r^{(1)} \mid 0 \leq r \leq m^{(1)} \right\}$  and
2. if  $n^{(1)} \geq 1$  then  $\delta^{v(i'_0{}^{(1)}, i'_1{}^{(1)})}$  leads to a  $v(i'_r{}^{(1)}, i'_{r+1}{}^{(1)})$ -loop for all  $1 \leq r < n^{(1)}$ .

We start over after position  $i'_1{}^{(1)}$  and set  $m^{(2)} =_{\text{def}} n^{(1)} - 1$ ,  $\Delta^{(2)} =_{\text{def}} \Delta^{(1)} \cup \left\{ \delta^{v(i'_0{}^{(1)}, i'_1{}^{(1)})} \right\}$  and  $i_r^{(2)} =_{\text{def}} i'_{r+1}{}^{(1)}$  for  $0 \leq r < n^{(1)}$ . We apply Lemma 1.12 again.

In general, after the  $j$ -th application of Lemma 1.12, we obtain for  $n^{(j)} =_{\text{def}} \lfloor m^{(j)}/n^n \rfloor$  the indices  $i'_r{}^{(j)}$  with  $0 \leq r \leq n^{(j)}$  such that

1.  $\left\{ i'_r{}^{(j)} \mid 0 \leq r \leq n^{(j)} \right\} \subseteq \left\{ i_r^{(j)} \mid 0 \leq r \leq m^{(j)} \right\}$ ,
2. if  $n^{(j)} \geq 1$  then every  $\delta' \in \Delta^{(j)}$  leads to a  $v(i'_0{}^{(j)}, i'_1{}^{(j)})$ -loop, and
3. if  $n^{(j)} \geq 1$  then every  $\delta' \in \Delta^{(j)} \cup \left\{ \delta^{v(i'_0{}^{(j)}, i'_1{}^{(j)})} \right\}$  leads to a  $v(i'_r{}^{(j)}, i'_{r+1}{}^{(j)})$ -loop for all  $1 \leq r < n^{(j)}$ .

Moreover, with  $m^{(j+1)} =_{\text{def}} n^{(j)} - 1$ ,  $\Delta^{(j+1)} =_{\text{def}} \Delta^{(j)} \cup \left\{ \delta^{v(i'_0{}^{(j)}, i'_1{}^{(j)})} \right\}$  and  $i_r^{(j+1)} =_{\text{def}} i'_{r+1}{}^{(j)}$  for  $0 \leq r < n^{(j)}$  we can carry out the  $(j+1)$ -st application of Lemma 1.12.

We chose  $l$  at the beginning large enough such that we can apply Lemma 1.12 sufficiently often to face the same mapping twice. This can be seen as follows. By Proposition 1.11

we have that  $m^{(j)} \geq (2n^n)^{(n^n+3-j)}$  for  $1 \leq j \leq n^n + 1$ . It follows that  $n^{(j)} = \lfloor m^{(j)}/n^n \rfloor \geq (2n^n)^{(n^n+2-j)} - 1 \geq 1$  for  $1 \leq j \leq n^n + 1$ . Particularly, the indices  $i'_0^{(j)}$  and  $i'_1^{(j)}$  exist for  $1 \leq j \leq n^n + 1$ .

On the one hand, at the end of each step  $j$  we take  $\delta^{v(i'_0^{(j)}, i'_1^{(j)})}$  to  $\Delta^{(j)}$  and obtain  $\Delta^{(j+1)}$ . On the other hand, there are at most  $n^n$  total mappings  $S \rightarrow S$ . Therefore, there exists a step  $t$  with  $1 \leq t \leq n^n + 1$  such that  $\bar{\delta} =_{\text{def}} \delta^{v(i'_0^{(t)}, i'_1^{(t)})}$  is already an element of  $\Delta^{(t)}$ . From the second statement of Lemma 1.12 it follows that  $\bar{\delta}$  leads to a  $v(i'_0^{(t)}, i'_1^{(t)})$ -loop. With  $g =_{\text{def}} i'_0^{(t)}$ ,  $h =_{\text{def}} i'_1^{(t)}$  and  $u =_{\text{def}} v_g v_{g+1} \cdots v_{h-1}$  we have  $u = v(i'_0^{(t)}, i'_1^{(t)})$ . Thus  $\bar{\delta} = \delta^u$  leads to a  $u$ -loop and hence  $\delta^{uu} = \delta^u$ .  $\square$

Now we are able to prove Theorem 1.14. If we do not have the particular number  $l$  of words, but factors  $v_0, v_1, \dots, v_n$  for arbitrary  $n$ , then we can partition them in a number of factors such that in each factor there are only  $\mathcal{I}_{\mathcal{M}}$  words  $v_i$ , and every second factor  $u$  has in fact the property  $\delta^{uu} = \delta^u$ .

**Theorem 1.14.** *For every DFA  $\mathcal{M} = (A, S, \delta, s_0, S')$  and for all  $v_0, \dots, v_n \in A^+$  there exist an  $m \geq 0$  and indices  $0 = i_0 < i_1 < \cdots < i_{2m+1} = n + 1$  such that*

1.  $i_{j+1} - i_j \leq \mathcal{I}_{\mathcal{M}}$  for  $0 \leq j \leq 2m$  and
2.  $\delta^{uu} = \delta^u$  for all  $u = v_{i_j} v_{i_j+1} \cdots v_{i_{j+1}-1}$  with  $1 \leq j < 2m$  and  $j \equiv 1 \pmod{2}$ .

*Proof.* Let  $l =_{\text{def}} \lfloor \mathcal{I}_{\mathcal{M}}/2 \rfloor$ . If  $n < \mathcal{I}_{\mathcal{M}}$  then we set  $m =_{\text{def}} 0$ ,  $i_0 =_{\text{def}} 0$ ,  $i_1 =_{\text{def}} n + 1$  and we are done. Otherwise, we disregard  $v_0$  and partition the list  $v_1, \dots, v_n$  from left to right into factors such that every factor contains  $l + 1$  words  $v_j$ . We obtain  $m \geq 1$  such factors  $B_1, \dots, B_m$  and  $r \leq l$  remaining words  $v_{n-r+1}, \dots, v_n$ . For every factor  $B_t = (v_j, v_{j+1}, \dots, v_{j+l})$  with  $j = (t-1)(l+1) + 1$  and  $1 \leq t \leq m$  we apply Lemma 1.13 and we obtain indices  $j \leq g_t < h_t \leq j + l$  such that  $\delta^{uu} = \delta^u$  with  $u =_{\text{def}} v_{g_t} v_{g_t+1} \cdots v_{h_t-1}$ . Now let  $i_0 =_{\text{def}} 0$ ,  $i_{2m+1} =_{\text{def}} n + 1$  and  $i_{2t-1} =_{\text{def}} g_t$ ,  $i_{2t} =_{\text{def}} h_t$  for  $1 \leq t \leq m$ . Observe that  $0 = i_0 < i_1 < \cdots < i_{2m+1} = n + 1$ . Moreover, we already have the second statement of Theorem 1.14.

It remains to show the first statement. For  $1 \leq t \leq m$  it holds that  $i_{2t} - i_{2t-1} = h_t - g_t \leq l < \mathcal{I}_{\mathcal{M}}$ . For  $1 \leq t < m$  we have  $B_t = (v_j, v_{j+1}, \dots, v_{j+l})$  and  $B_{t+1} = (v_{j+l+1}, v_{j+l+2}, \dots, v_{j+2l+1})$  with  $j = (t-1)(l+1) + 1$ . Since

$$j \leq g_t < h_t \leq j + l < j + l + 1 \leq g_{t+1} < h_{t+1} \leq j + 2l + 1$$

it follows that  $g_{t+1} - h_t \leq (j + 2l) - (j + 1) = 2l - 1 < \mathcal{I}_{\mathcal{M}}$ . Moreover  $i_1 - i_0 = g_1 \leq l < \mathcal{I}_{\mathcal{M}}$ , so we have shown  $i_{j+1} - i_j \leq \mathcal{I}_{\mathcal{M}}$  for  $0 \leq j \leq 2m - 1$ .

We are left with  $i_{2m+1} - i_{2m}$ . Observe that  $B_m = (v_{n-r-l}, v_{n-r-l+1}, \dots, v_{n-r})$  and that  $i_{2m} = h_m > n - r - l$ . So

$$i_{2m+1} - i_{2m} = n + 1 - i_{2m} < n + 1 - n + r + l = r + l + 1 \leq 2l + 1 \leq \mathcal{I}_{\mathcal{M}} + 1$$

and hence  $i_{2m+1} - i_{2m} \leq \mathcal{I}_{\mathcal{M}}$ .  $\square$

**Corollary 1.15.** *For every DFA  $\mathcal{M} = (A, S, \delta, s_0, S')$  and every  $w \in A^+$  there exist words  $w_0, \dots, w_m, u_1, \dots, u_m \in A^{+\leq \mathcal{I}_{\mathcal{M}}}$  with  $w = w_0 u_1 w_1 \cdots u_m w_m$  and  $\delta^{u_i} = \delta^{u_i u_i}$  for  $1 \leq i \leq m$ .*

*Proof.* This follows from Theorem 1.14 when considering letters  $v_0, \dots, v_n \in A$  such that  $w = v_0 \cdots v_n$ .  $\square$

**Corollary 1.16.** *For every DFA  $\mathcal{M} = (A, S, \delta, s_0, S')$  and for all  $v_1, \dots, v_n \in A^+$  with  $n \geq \mathcal{I}_{\mathcal{M}}$  there exist  $i, j$  with  $1 \leq i \leq j \leq n$  and  $j - i < \mathcal{I}_{\mathcal{M}}$  such that  $\delta^{u u} = \delta^u$  for  $u =_{\text{def}} v_i v_{i+1} \cdots v_j$ .*

*Proof.* Let  $v_0 =_{\text{def}} a$  for some letter  $a$ . If we apply Theorem 1.14 to  $v_0, \dots, v_n$  then we obtain an  $m \geq 0$  and indices  $0 = i_0 < i_1 < \cdots < i_{2m+1} = n + 1$  such that

1.  $i_{j+1} - i_j \leq \mathcal{I}_{\mathcal{M}}$  for  $0 \leq j \leq 2m$  and
2.  $\delta^{u u} = \delta^u$  for all  $u = v_{i_j} v_{i_j+1} \cdots v_{i_{j+1}-1}$  with  $1 \leq j < 2m$  and  $j \equiv 1 \pmod{2}$ .

If  $m = 0$  then it follows that  $i_1 = n + 1$  and  $i_1 - i_0 = n + 1 > \mathcal{I}_{\mathcal{M}}$ . This contradicts the first statement and it follows that  $m \geq 1$ . In particular, there exist indices  $0 = i_0 < i_1 < i_2 < i_3 \leq n + 1$  with  $i_2 - i_1 \leq \mathcal{I}_{\mathcal{M}}$  and  $\delta^{u u} = \delta^u$  for  $u =_{\text{def}} v_{i_1} v_{i_1+1} \cdots v_{i_2-1}$ . Therefore, with  $i =_{\text{def}} i_1$  and  $j =_{\text{def}} i_2 - 1$  we have found indices with  $1 \leq i \leq j \leq n$  and  $j - i = i_2 - 1 - i_1 < \mathcal{I}_{\mathcal{M}}$  such that  $\delta^{u u} = \delta^u$  for  $u = v_i v_{i+1} \cdots v_j$ .  $\square$

**Corollary 1.17.** *For every DFA  $\mathcal{M} = (A, S, \delta, s_0, S')$  and every  $w \in A^{*\geq \mathcal{I}_{\mathcal{M}}}$  there exist words  $u, w_1, w_2 \in A^*$  such that  $1 \leq |u| \leq \mathcal{I}_{\mathcal{M}}$ ,  $\delta^u = \delta^{u u}$  and  $w = w_1 u w_2$ .*

*Proof.* This follows from Corollary 1.16 when considering letters  $v_1, \dots, v_n \in A$  such that  $w = v_1 \cdots v_n$ .  $\square$

## 1.4 $\prec_v^{0,k}$ and $\prec_v^{1,k}$ Word Extensions

In our approach we use word extensions to prove decidability results for concatenation hierarchies. These word extensions can be also considered as binary relations on the set of words. Roughly speaking, they are such that a certain factor is inserted at a certain position in an initial word. Here the emphasis is on “certain position” which means a position where a special word—a so-called context word—appears. So our word extensions are determined by two things: (i) by the “certain position” where an extension is allowed, i.e., by the possible context words, and (ii) by the “certain factor” that is inserted.

In this section we introduce the word extensions  $\prec_v^{0,k}$  and  $\prec_v^{1,k}$ . They should be considered as elementary extensions, since in chapter 4 we will use them as building blocks to define more complicated extensions. More precisely,  $\prec_v^{0,k}$  (respectively,  $\prec_v^{1,k}$ ) will be used to define the extensions  $\preceq_{\mathcal{M}}^{0,k}$  (respectively,  $\preceq_{\mathcal{M}}^{1,k}$ ) which in turn are used to prove the decidability of the levels 1/2 (respectively, the levels 3/2) of the dot-depth hierarchy and of the Straubing-Thérien hierarchy (see chapter 2).

After proving some basic properties of  $\prec_v^{0,k}$  and  $\prec_v^{1,k}$  we introduce classes of languages  $\tilde{\mathcal{B}}_k$  for  $k \geq 0$ . From their definition it will be easy to see that  $\tilde{\mathcal{B}}_0$  is contained in level 3/2

of the Straubing-Thérien hierarchy, and that  $\tilde{\mathcal{B}}_k$  is contained in level 3/2 of the dot-depth hierarchy for all  $k \geq 0$ . The main result in this section is Theorem 1.30 which says that the  $\langle \cdot \rangle_v^{1,k}$  upward closure of a language from  $\tilde{\mathcal{B}}_k$  is in  $\tilde{\mathcal{B}}_k$  again.

### 1.4.1 Basic Definitions

We start with the definition of  $\langle \cdot \rangle_v^{0,k}$  and  $\langle \cdot \rangle_v^{1,k}$  and prove some basic properties of these word extensions.

**Definition 1.18.** *Let  $k \geq 0$  and  $v, w, w' \in A^*$ .*

$$\begin{aligned} w \langle \cdot \rangle_v^{0,k} w' &\iff_{\text{def}} \text{there exist words } x, z \in A^* \text{ and } u \in A^{\geq k+1} \text{ such that} \\ &\quad w = xvz \text{ and } w' = xvuvz \\ w \langle \cdot \rangle_v^{1,k} w' &\iff_{\text{def}} \text{there exist words } x, z \in A^* \text{ and } u \in A^{\geq k+1} \text{ such that} \\ &\quad w = xvz, w' = xvuvz \text{ and } \alpha_k(vuv) \subseteq \alpha_k(v) \end{aligned}$$

Note that for  $\langle \cdot \rangle_v^{1,k}$  we always have  $|v| \geq k + 1$  because  $\alpha_k(vuv) \subseteq \alpha_k(v)$  and  $|u| \geq k + 1$ . Moreover,  $\langle \cdot \rangle_v^{0,k}$  and  $\langle \cdot \rangle_v^{1,k}$  are not reflexive and not transitive. However, in chapter 4 we will introduce reflexive and transitive word extensions  $\preceq_{\mathcal{M}}^{0,k}$  and  $\preceq_{\mathcal{M}}^{1,k}$  which are sequences of  $\langle \cdot \rangle_v^{0,k}$  and  $\langle \cdot \rangle_v^{1,k}$  extensions. We state the following easy facts about  $\langle \cdot \rangle_v^{0,k}$  and  $\langle \cdot \rangle_v^{1,k}$  which says in particular they are stable.

**Proposition 1.19.** *For  $k \geq 0$  and  $v, w, w' \in A^*$  the following holds.*

1. *If  $w \langle \cdot \rangle_v^{0,k} w'$  or  $w \langle \cdot \rangle_v^{1,k} w'$  then  $\mathfrak{p}_{|v|}(w) = \mathfrak{p}_{|v|}(w')$  and  $\mathfrak{s}_{|v|}(w) = \mathfrak{s}_{|v|}(w')$ .*
2. *If  $w \langle \cdot \rangle_v^{0,k} w'$  then  $xwz \langle \cdot \rangle_v^{0,k} xw'z$  for all  $x, z \in A^*$ .*
3. *If  $w \langle \cdot \rangle_v^{1,k} w'$  then  $xwz \langle \cdot \rangle_v^{1,k} xw'z$  for all  $x, z \in A^*$ .*
4. *If  $w \langle \cdot \rangle_v^{1,k} w'$  then  $\alpha_k(w) = \alpha_k(w')$ .*

*Proof.* These are easy consequences of Definition 1.18 (note that  $|v| \geq k + 1$  in statement 4).  $\square$

**Proposition 1.20.** *For  $k \geq 0$ ,  $v, w \in A^*$  and  $w' \in \langle w \rangle_{\langle \cdot \rangle_v^{1,k}}$  the following holds.*

1.  $\mathfrak{p}_{|v|}(w) = \mathfrak{p}_{|v|}(w')$  and  $\mathfrak{s}_{|v|}(w) = \mathfrak{s}_{|v|}(w')$
2.  $\langle xw \rangle_{\langle \cdot \rangle_v^{1,k}} \subseteq xA^{\geq |w|}$  and  $\langle wz \rangle_{\langle \cdot \rangle_v^{1,k}} \subseteq A^{\geq |w|}z$  for  $x, z \in A^{|v|}$
3.  $\langle xwz \rangle_{\langle \cdot \rangle_v^{1,k}} \subseteq xA^{\geq |w|}z$  for  $x, z \in A^{|v|}$
4.  $xw'z \in \langle xwz \rangle_{\langle \cdot \rangle_v^{1,k}}$  for  $x, z \in A^*$
5.  $\alpha_k(w) = \alpha_k(w')$
6.  $\langle L_1 \rangle_{\langle \cdot \rangle_v^{1,k}} \cdot \langle L_2 \rangle_{\langle \cdot \rangle_v^{1,k}} \cdots \langle L_n \rangle_{\langle \cdot \rangle_v^{1,k}} \subseteq \langle L_1 L_2 \cdots L_n \rangle_{\langle \cdot \rangle_v^{1,k}}$  for  $n \geq 1$  and  $L_1, \dots, L_n \subseteq A^*$

*Proof.* This follows from Proposition 1.19 (note that  $w = w'$  if  $|v| \leq k$ ).  $\square$

We introduce classes of languages  $\tilde{\mathcal{B}}_k$ . Later we will see that they are contained in the level 3/2 of the dot-depth hierarchy.

**Definition 1.21.** For  $k \geq 0$ ,  $\beta, \delta \in A^k$  and  $\Gamma \subseteq A^{k+1}$  we define the following.

$$\begin{aligned} (\beta|\Gamma|\delta)_k &=_{\text{def}} \left\{ w \in A^{*\geq k+1} \mid \mathfrak{p}_k(w) = \beta, \mathfrak{s}_k(w) = \delta, \alpha_k(w) \subseteq \Gamma \right\} \\ \tilde{\mathcal{B}}_k &=_{\text{def}} \text{Pol} \left( \left\{ (\beta|\Gamma|\delta)_k \mid \beta, \delta \in A^k, \Gamma \subseteq A^{k+1} \right\} \cup \{ \{a\} \mid a \in A \} \right) \end{aligned}$$

From Definition 1.18 we see that in  $\langle_v^{1,k}$  extensions not all words  $u$  can be used for insertions but only words  $u \in A^{*\geq k+1}$  with  $\alpha_k(vuv) \subseteq \alpha_k(v)$ . In the proofs we will often need these words, and therefore we make the following definition.

**Definition 1.22.** For  $k \geq 0$  and  $v \in A^{*\geq k+1}$  let

$$L_{k,v} =_{\text{def}} \bigcup_{\substack{\beta, \delta \in A^k \text{ such that} \\ \alpha_k(v\beta), \alpha_k(\delta v) \subseteq \alpha_k(v)}} (\beta|\alpha_k(v)|\delta)_k.$$

Note that  $L_{k,v} \in \tilde{\mathcal{B}}_k$  for  $k \geq 0$  and  $v \in A^{*\geq k+1}$ .

**Proposition 1.23.** For  $k \geq 0$  and  $v \in A^{*\geq k+1}$  it holds that

$$L_{k,v} = \left\{ w \in A^{*\geq k+1} \mid \alpha_k(vwv) = \alpha_k(v) \right\} = \left\{ w \in A^{*\geq k+1} \mid \alpha_k(vwv) \subseteq \alpha_k(v) \right\}.$$

*Proof.* This follows from Definition 1.22.  $\square$

This means that  $L_{k,v}$  is exactly the set of words  $u$  that can be inserted by  $\langle_v^{1,k}$  extensions (see Definition 1.18). In particular it holds that  $xvz \langle_v^{1,k} xvuvz$  for all  $k \geq 0$ ,  $v \in A^{*\geq k+1}$ ,  $x, z \in A^*$  and  $u \in L_{k,v}$ .

In chapter 4 we will investigate the upward closure of a word  $y \in A^+$  under word extensions  $\langle_{v_1}^{1,k}, \langle_{v_2}^{1,k}, \dots, \langle_{v_n}^{1,k}$ . This means the set of words that can be reached from  $y$  by the (repeated) use of these extensions for certain words  $v_1, v_2, \dots, v_n$ . There we will need to show that this set is contained in  $\tilde{\mathcal{B}}_k$  (and therefore in level 3/2 of the dot-depth hierarchy). To prepare this result we show in the following subsections that the  $\langle_v^{1,k}$  upward closure of languages from  $\tilde{\mathcal{B}}_k$  is in  $\tilde{\mathcal{B}}_k$  again. In order to do this we refine the class  $\tilde{\mathcal{B}}_k$  in the following subsection.

### 1.4.2 The Classes $\tilde{\mathcal{B}}_{k,m}$

For  $k \geq 0$  we introduce classes  $\tilde{\mathcal{B}}_{k,m}$  that refine the class  $\tilde{\mathcal{B}}_k$ . We show that all classes  $\tilde{\mathcal{B}}_{k,m}$  are closed under concatenation with words and under intersection with  $A^*w$  and  $wA^*$  for  $w \in A^*$ . The latter means that we can test for certain prefixes and suffixes.

**Definition 1.24.** Let  $k \geq 0$  and  $m \geq 1$ . Define  $\tilde{\mathcal{B}}_{k,m}$  as the class of finite (possibly empty) unions of languages  $L = L_1L_2 \cdots L_n \subseteq A^+$  with  $1 \leq n \leq m$  and for  $1 \leq i \leq n$  it holds that

$$\begin{aligned} L_i &= \{w_i\} \quad \text{for } w_i \in A^+, \text{ or} \\ L_i &= x_i(\beta_i|\Gamma_i|\delta_i)_k z_i \quad \text{for } x_i, z_i \in A^*, \beta_i, \delta_i \in A^k \text{ and } \Gamma_i \subseteq A^{k+1}. \end{aligned}$$

As an easy consequence of this definition we obtain that the union of the classes  $\tilde{\mathcal{B}}_{k,m}$  over  $m \geq 1$  is equal to  $\tilde{\mathcal{B}}_k$ . Moreover,  $\tilde{\mathcal{B}}_{k,m}$  has the following closure properties.

**Proposition 1.25.** *Let  $k \geq 0$ ,  $m \geq 1$ ,  $L \in \tilde{\mathcal{B}}_{k,m}$ , and  $w \in A^*$ . Then it holds that*

1.  $Lw \in \tilde{\mathcal{B}}_{k,m}$ ,  $wL \in \tilde{\mathcal{B}}_{k,m}$ ,
2.  $L \cap A^*w \in \tilde{\mathcal{B}}_{k,m}$ , and  $L \cap wA^* \in \tilde{\mathcal{B}}_{k,m}$ .

*Proof.* Statement 1 follows immediately from the definition of  $\tilde{\mathcal{B}}_{k,m}$ . For the second statement we need the following claim which can be easily verified.

**Claim.** *Let  $n =_{\text{def}} \max\{|w|, k\}$ ,  $\beta, \delta \in A^k$  and  $\Gamma \subseteq A^{k+1}$ . Then it holds that*

$$\begin{aligned} (\beta|\Gamma|\delta)_k &= \left( \underbrace{\bigcup_{\substack{\beta', \delta' \in A^k, x \in \beta A^*, z \in A^* \delta \text{ such that} \\ |x|=|z|=n, \alpha_k(x\beta') \subseteq \Gamma, \alpha_k(\delta'z) \subseteq \Gamma}} x \cdot (\beta'|\Gamma|\delta')_k \cdot z}_{= (\beta|\Gamma|\delta)_k \cap A^{\geq k+1+2n}} \right) \cup \left( \underbrace{\bigcup_{\substack{w' \in (\beta|\Gamma|\delta)_k \text{ with} \\ |w'| \leq k+2n}} \{w'\}}_{= (\beta|\Gamma|\delta)_k \cap A^{\leq k+2n}} \right). \end{aligned}$$

From this claim it follows that  $L$  can be written as a finite union of languages  $L' = L_1L_2 \cdots L_j$  such that  $1 \leq j \leq m$  and for  $1 \leq i \leq j$  it holds that either  $L_i = \{w_i\}$  for some  $w_i \in A^+$ , or  $L_i = x_i(\beta_i|\Gamma_i|\delta_i)_k z_i$  for  $x_i, z_i \in A^{\geq |w|}$ ,  $\beta_i, \delta_i \in A^k$  and  $\Gamma_i \subseteq A^{k+1}$ . It suffices to show that  $L_1L_2 \cdots L_j \cap A^*w \in \tilde{\mathcal{B}}_{k,m}$ .

**Case 1:** Assume that  $L_i = \{w_i\}$  for all  $1 \leq i \leq j$ . Then  $L_1L_2 \cdots L_j = \{w'\}$  for some  $w' \in A^+$ , and it follows that either  $L_1L_2 \cdots L_j \cap A^*w = \{w'\}$  or  $L_1L_2 \cdots L_j \cap A^*w = \emptyset$ . So in this case we have  $L_1L_2 \cdots L_j \cap A^*w \in \tilde{\mathcal{B}}_{k,m}$ .

**Case 2:** Assume that not all  $L_i$  are of the form  $L_i = \{w_i\}$ . Then there exists a maximal  $l$  with  $1 \leq l \leq j$  such that  $L_l = x_l(\beta_l|\Gamma_l|\delta_l)_k z_l$  and  $L_i = \{w_i\}$  for all  $l < i \leq j$ . In this case it holds that  $L_1L_2 \cdots L_j = L_1L_2 \cdots L_{l-1}x_l(\beta_l|\Gamma_l|\delta_l)_k z_l w'$  where  $w' =_{\text{def}} w_{l+1}w_{l+2} \cdots w_j$ . Since  $|z_l| \geq |w|$  we obtain

$$L_1L_2 \cdots L_j \cap A^*w = \begin{cases} L_1L_2 \cdots L_j & : \text{ if } w \text{ is a suffix of } z_l w' \\ \emptyset & : \text{ otherwise.} \end{cases}$$

So also in this case we obtain  $L_1L_2 \cdots L_j \cap A^*w \in \tilde{\mathcal{B}}_{k,m}$ . It follows that  $L \cap A^*w \in \tilde{\mathcal{B}}_{k,m}$ , and analogously we show  $L \cap wA^* \in \tilde{\mathcal{B}}_{k,m}$ .  $\square$

### 1.4.3 The $\langle_v^{1,k}$ Upward Closure of a Word

We want to show that the  $\langle_v^{1,k}$  upward closure of a nonempty word is in  $\tilde{\mathcal{B}}_k$ . The idea is as follows: Let  $y$  be a word and let  $y' \in \langle y \rangle_{\langle_v^{1,k}}$ . This means that  $y'$  emerges from  $y$  by a sequence of  $\langle_v^{1,k}$  extensions. By definition, a single  $\langle_v^{1,k}$  extension is such that a given word is modified by inserting some letters at exactly one position in this word. With the following lemma we show that in the sequence leading from  $y$  to  $y'$  one can trace back these positions. This yields a list of positions in  $y$  that can be used to transform  $y$  into  $y'$  in a single step (where  $\langle_v^{1,k}$  extensions are carried out, in parallel, at several positions in  $y$ ). The fact that the number of these positions is  $\leq |y| + 1$  means that there exists a sequence of length  $\leq |y| + 1$  leading from  $y$  to  $y'$ . This implies  $\langle y \rangle_{\langle_v^{1,k}} \in \tilde{\mathcal{B}}_k$ .

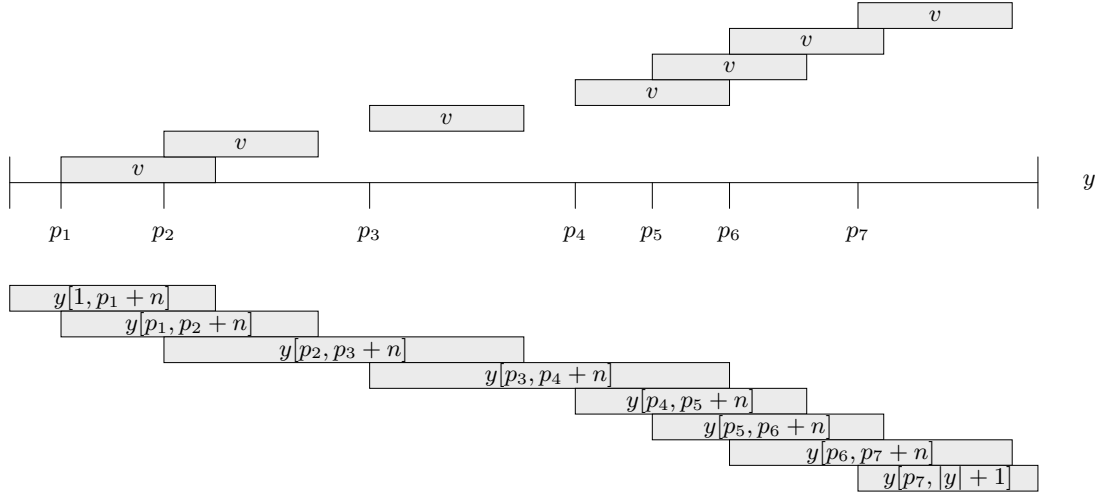


**Lemma 1.26.** For  $k \geq 0$ ,  $y \in A^*$ ,  $v \in A^{*\geq k+1}$  and  $n =_{\text{def}} |v|$  it holds that

$$\langle y \rangle_{\prec_v^{1,k}} = \{y\} \cup \bigcup y[1, p_1 + n] \cdot L_{k,v} \cdot y[p_1, p_2 + n] \cdot L_{k,v} \cdot y[p_2, p_3 + n] \cdot \cdots \cdot L_{k,v} \cdot y[p_{m-1}, p_m + n] \cdot L_{k,v} \cdot y[p_m, |y| + 1] \quad (1.1)$$

where the union ranges over all  $m \geq 1$  and all positions  $1 \leq p_1 < p_2 < \cdots < p_m \leq |y| - n + 1$  with  $y[p_i, p_i + n] = v$  for  $1 \leq i \leq m$ .

*Proof.* The idea behind the union above is illustrated in the following picture. It shows the factors that emerge when we consider the positions  $p_1, p_2, \dots, p_7$  in  $y$ . The upper part of the picture shows the  $v$ -blocks that appear at the positions  $p_i$  and that have to be doubled when making  $\prec_v^{1,k}$  extensions at  $p_i$ . In the lower part we see the factors of  $y$  that remain connected. Note that in the lower part, neighboring factors overlap in exactly  $n = |v|$  letters and all these overlapping parts are equal to  $v$ .



In the proof we denote the right-hand side of (1.1) by  $L$ . At first we show  $\langle y \rangle_{\prec_v^{1,k}} \subseteq L$ . For this we assume that  $\langle y \rangle_{\prec_v^{1,k}} \not\subseteq L$ , this will lead to a contradiction. Since at least  $y$  is in  $L$ , there exist words  $w, w' \in A^*$  such that  $w \in L$ ,  $w' \notin L$  and  $w \prec_v^{1,k} w'$ . Hence there exist words  $x, z \in A^*$  and  $u \in L_{k,v}$  such that  $w = xvz$  and  $w' = xvuvz$ .

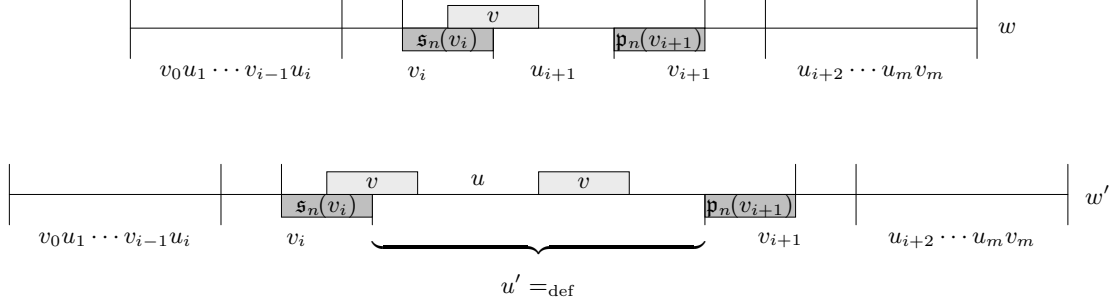
If  $w = y$  then with  $p_1 =_{\text{def}} |x| + 1$  we obtain  $y[1, p_1 + n] = xv$  and  $y[p_1, |y| + 1] = vz$ . Therefore, we get  $w' \in y[1, p_1 + n] \cdot L_{k,v} \cdot y[p_1, |y| + 1] \subseteq L$  which contradicts our assumption.

Assume now  $w \neq y$ . Then there exist an  $m \geq 1$ , positions  $1 \leq p_1 < \cdots < p_m \leq |y| - n + 1$  and words  $u_1, \dots, u_m \in L_{k,v}$  such that

$$w = \underbrace{y[1, p_1 + n]}_{v_0 = \text{def}} \cdot u_1 \cdot \underbrace{y[p_1, p_2 + n]}_{v_1 = \text{def}} \cdot u_2 \cdot \underbrace{y[p_2, p_3 + n]}_{v_2 = \text{def}} \cdot \cdots \cdot u_m \cdot \underbrace{y[p_m, |y| + 1]}_{v_m = \text{def}}. \quad (1.2)$$

For  $0 \leq i \leq m$  define  $v_i$  as above and note that  $|v_i| \geq n$  (for  $1 \leq i \leq m - 1$  it even holds that  $|v_i| \geq n + 1$ ). Now we compare the decompositions (1.2) and  $w = xvz$ .

**Case 1:** Assume that the factor  $v$  of the decomposition  $w = x v z$  is contained in some factor  $\mathfrak{s}_n(v_i) \cdot u_{i+1} \cdot \mathfrak{p}_n(v_{i+1})$  of the decomposition (1.2). Then we have the following situation.



Define  $u'$  as in the picture above and note that  $w' = v_0 u_1 v_1 \dots u_i v_i u' v_{i+1} \dots u_m v_m$ . Moreover, it holds that  $\mathfrak{s}_n(v_i) u_{i+1} \mathfrak{p}_n(v_{i+1}) <_{v}^{1,k} \mathfrak{s}_n(v_i) u' \mathfrak{p}_n(v_{i+1})$ , and therefore  $v u_{i+1} v <_{v}^{1,k} v u' v$ . From Proposition 1.19.4 it follows that  $\alpha_k(v u_{i+1} v) = \alpha_k(v u' v)$ . By Proposition 1.23 we have  $\alpha_k(v u_{i+1} v) = \alpha_k(v)$  because  $u_{i+1} \in L_{k,v}$ . Therefore,  $\alpha_k(v u' v) = \alpha_k(v)$  and it follows that  $u' \in L_{k,v}$  (note that  $|u'| \geq k + 1$ ). This shows

$$w' \in y[1, p_1 + n] \cdot L_{k,v} \cdot y[p_1, p_2 + n] \cdot L_{k,v} \cdot y[p_2, p_3 + n] \cdots L_{k,v} \cdot y[p_m, |y| + 1]$$

and we get a contradiction to the assumption  $w' \notin L$ .

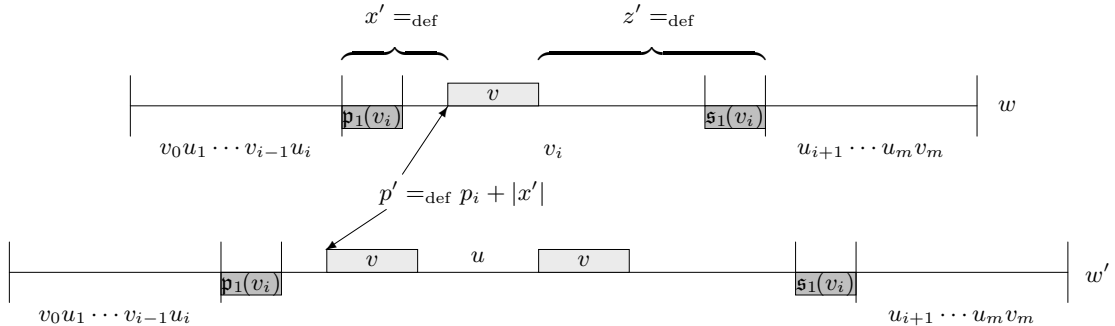
**Case 2:** Assume now that the factor  $v$  of the decomposition  $w = x v z$  is *not* contained in some factor  $\mathfrak{s}_n(v_i) \cdot u_{i+1} \cdot \mathfrak{p}_n(v_{i+1})$  of the decomposition (1.2). Since all  $\mathfrak{s}_n(v_i) \cdot u_{i+1} \cdot \mathfrak{p}_n(v_{i+1})$  are of length  $\geq n = |v|$ , it must be that one of the following subcases occurs.

*Case 2a:*  $v$  is contained in some factor  $A^{-1} v_i A^{-1}$  for  $1 \leq i \leq m - 1$

*Case 2b:*  $v$  is contained in  $v_0 A^{-1}$

*Case 2c:*  $v$  is contained in  $A^{-1} v_m$

We will only treat the Case 2a, the other cases are analogous. Hence our current situation is as follows.



Define  $p'$  as in the picture. Since  $x', z'$  are nonempty we have  $1 \leq |x'| \leq |v_i| - n - 1$ . Together with  $|v_i| = p_{i+1} - p_i + n$  this implies  $1 \leq |x'| \leq p_{i+1} - p_i - 1$ , and therefore  $p_i < p' < p_{i+1}$ . We obtain  $y[p', p' + n] = v$  because  $v_i = y[p_i, p_{i+1} + n]$  and  $v_i[|x'| + 1, |x'| + 1 + n] = v$ . Now consider the term of the union in  $L$  that takes the positions

$$1 \leq p_1 < \cdots < p_i < p' < p_{i+1} < \cdots < p_m \leq |y| - n + 1$$

into account. Since  $y[p_i, p' + n] = x'v$  and  $y[p', p_{i+1} + n] = vz'$  this term is equal to

$$\begin{aligned} L' =_{\text{def}} & v_0 \cdot L_{k,v} \cdot v_1 \cdot L_{k,v} \cdots v_{i-1} \cdot L_{k,v} \cdot x'v \cdot L_{k,v} \cdot vz' \cdot L_{k,v} \cdot v_{i+1} \\ & \cdot L_{k,v} \cdot v_{i+2} \cdots L_{k,v} \cdot v_{m-1} \cdot L_{k,v} \cdot v_m. \end{aligned}$$

Since  $w' = v_0 u_1 v_1 u_2 \cdots v_{i-1} u_i \cdot x'v \cdot u \cdot vz' \cdot u_{i+1} v_{i+1} \cdots u_{m-1} v_{m-1} u_m v_m$  and since  $u \in L_{k,v}$  we get  $w' \in L' \subseteq L$  which contradicts our assumption.

So in all considered cases we get contradictions. Therefore, our assumption was false and it follows that  $\langle y \rangle_{\langle v \rangle}^{1,k} \subseteq L$ . So we have shown that the left-hand side is a subset of the right-hand side in (1.1).

We turn to the proof of the reverse inclusion. Clearly, it holds that  $\langle y \rangle_{\langle v \rangle}^{1,k} \supseteq \{y\}$ . So let  $m \geq 1$  and choose positions  $1 \leq p_1 < \cdots < p_m \leq |y| - n + 1$  with  $y[p_i, p_i + n] = v$  for  $1 \leq i \leq m$ . It follows that

$$y[1, p_i] \cdot v = y[1, p_{i-1}] \cdot y[p_{i-1}, p_i + n] \quad \text{for } 2 \leq i \leq m. \quad (1.3)$$

In order to show that  $y' \in \langle y \rangle_{\langle v \rangle}^{1,k}$  for all  $y' \in L \setminus \{y\}$  we choose arbitrary  $u_1, \dots, u_m \in L_{k,v}$  and let

$$y' =_{\text{def}} y[1, p_1 + n] \cdot u_1 \cdot y[p_1, p_2 + n] \cdot u_2 \cdot y[p_2, p_3 + n] \cdots u_{m-1} \cdot y[p_{m-1}, p_m + n] \cdot u_m \cdot y[p_m, |y| + 1].$$

From Proposition 1.23 it follows that  $\alpha_k(vu_i v) = \alpha_k(v)$  for  $1 \leq i \leq m$ . Since  $y$  can be written as  $y = y[1, p_m] \cdot y[p_m, |y| + 1]$  and since  $y[p_m, |y| + 1]$  has the prefix  $v$ , we get  $y \langle_{\langle v \rangle}^{1,k} y_m$  for

$$y_m =_{\text{def}} y[1, p_m] \cdot v \cdot u_m \cdot y[p_m, |y| + 1].$$

By (1.3),  $y_m$  can also be written as  $y_m = y[1, p_{m-1}] \cdot y[p_{m-1}, p_m + n] \cdot u_m \cdot y[p_m, |y| + 1]$ . Since  $y[p_{m-1}, p_m + n]$  has the prefix  $v$  we obtain  $y_m \langle_{\langle v \rangle}^{1,k} y_{m-1}$  for

$$y_{m-1} =_{\text{def}} y[1, p_{m-1}] \cdot v \cdot u_{m-1} \cdot y[p_{m-1}, p_m + n] \cdot u_m \cdot y[p_m, |y| + 1].$$

We continue this argumentation until we obtain  $y_2 \langle_{\langle v \rangle}^{1,k} y_1$  for

$$y_1 =_{\text{def}} y[1, p_1] \cdot v \cdot u_1 \cdot y[p_1, p_2 + n] \cdot u_2 \cdot y[p_2, p_3 + n] \cdots u_{m-1} \cdot y[p_{m-1}, p_m + n] \cdot u_m \cdot y[p_m, |y| + 1].$$

Since  $y[1, p_1] \cdot v = y[1, p_1 + n]$  we have  $y_1 = y'$ , and therefore  $y \langle_{\langle v \rangle}^{1,k} y_m \langle_{\langle v \rangle}^{1,k} \cdots \langle_{\langle v \rangle}^{1,k} y_2 \langle_{\langle v \rangle}^{1,k} y'$ . This shows  $y' \in \langle y \rangle_{\langle v \rangle}^{1,k}$  and it follows that in (1.1) the right-hand side is a subset of the left-hand side.  $\square$

**Corollary 1.27.** For  $k \geq 0$ ,  $y \in A^+$  and  $v \in A^{*\geq k+1}$  it holds that  $\langle y \rangle_{\langle v \rangle}^{1,k} \in \tilde{\mathcal{B}}_k$ .

*Proof.* This follows from Lemma 1.26 and Definition 1.21 since the union in (1.1) is finite.  $\square$

#### 1.4.4 The $\langle \cdot \rangle_v^{1,k}$ Upward Closure of Languages from $\tilde{\mathcal{B}}_{k,1}$

We still prepare the proof of Theorem 1.30 where we will show that the  $\langle \cdot \rangle_v^{1,k}$  upward closure of a language from  $\tilde{\mathcal{B}}_k$  is in  $\tilde{\mathcal{B}}_k$  again. Since we already know that the classes  $\tilde{\mathcal{B}}_{k,m}$  exhaust the class  $\tilde{\mathcal{B}}_k$  we can prove this by induction on  $m \geq 1$ . In this subsection we show the corresponding induction base, i.e., we show that the  $\langle \cdot \rangle_v^{1,k}$  upward closure of languages from  $\tilde{\mathcal{B}}_{k,1}$  is in  $\tilde{\mathcal{B}}_k$ .

**Lemma 1.28.** *Let  $k \geq 0$ ,  $v \in A^{*\geq k+1}$  and  $L \in \tilde{\mathcal{B}}_{k,1}$ . Then it holds that  $\langle L \rangle_{\langle \cdot \rangle_v^{1,k}} \in \tilde{\mathcal{B}}_k$ .*

*Proof.* By Definition 1.24, it suffices to show the lemma for languages  $L$  that are either of the form  $L = \{w\}$  or of the form  $L = x(\beta|\Gamma|\delta)_k z$ . If  $L = \{w\}$  for some  $w \in A^+$  then  $\langle L \rangle_{\langle \cdot \rangle_v^{1,k}} \in \tilde{\mathcal{B}}_k$  by Corollary 1.27.

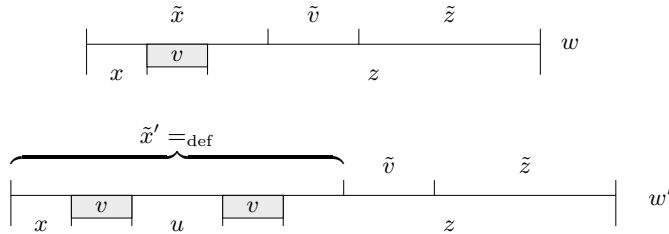
Assume now that  $L = x_1(\beta|\Gamma|\delta)_k z_1$  for  $\beta, \delta \in A^k$ ,  $\Gamma \subseteq A^{k+1}$  and  $x_1, z_1 \in A^*$ . By Corollary 1.27, it suffices to show  $\langle L \cap A^{*\geq n} \rangle_{\langle \cdot \rangle_v^{1,k}} \in \tilde{\mathcal{B}}_k$  where  $n =_{\text{def}} 2|v| + |x_1| + |z_1| + k + 1$ . So it is enough to prove that

$$\langle L \cap A^{*\geq n} \rangle_{\langle \cdot \rangle_v^{1,k}} = \bigcup_{\substack{x_2 \in A^{|v|} \cap \beta A^*, z_2 \in A^{|v|} \cap A^* \delta, \beta', \delta' \in A^k \\ \text{such that } \alpha_k(x_2 \beta') \subseteq \Gamma \text{ and } \alpha_k(\delta' z_2) \subseteq \Gamma}} \langle x_1 x_2 \rangle_{\langle \cdot \rangle_v^{1,k}} \cdot (\beta'|\Gamma|\delta')_k \cdot \langle z_2 z_1 \rangle_{\langle \cdot \rangle_v^{1,k}}. \quad (1.4)$$

To see this inclusion from right to left we observe that  $x_1 x_2 \cdot (\beta'|\Gamma|\delta')_k \cdot z_2 z_1 \subseteq L \cap A^{*\geq n}$  for all  $x_2 \in A^{|v|} \cap \beta A^*$ ,  $z_2 \in A^{|v|} \cap A^* \delta$  and  $\beta', \delta' \in A^k$  with  $\alpha_k(x_2 \beta'), \alpha_k(\delta' z_2) \subseteq \Gamma$ . Together with Proposition 1.20.6 this implies  $\langle x_1 x_2 \rangle_{\langle \cdot \rangle_v^{1,k}} \cdot (\beta'|\Gamma|\delta')_k \cdot \langle z_2 z_1 \rangle_{\langle \cdot \rangle_v^{1,k}} \subseteq \langle L \cap A^{*\geq n} \rangle_{\langle \cdot \rangle_v^{1,k}}$ .

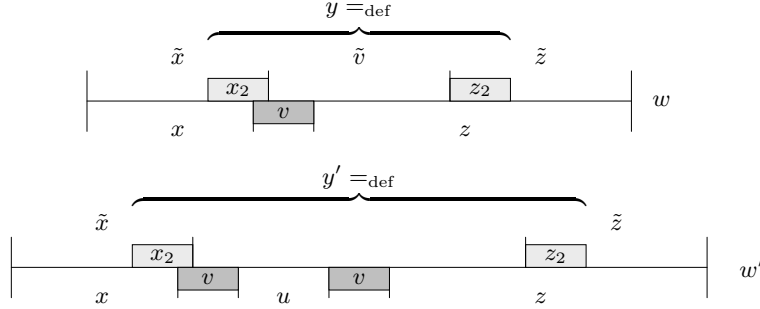
We turn to the other inclusion, i.e., the inclusion from left to right. We assume that the inclusion does not hold, this will lead to a contradiction. Observe that  $L \cap A^{*\geq n}$  is a subset of the right-hand side of (1.4). It follows that there exist words  $w, w' \in \langle L \cap A^{*\geq n} \rangle_{\langle \cdot \rangle_v^{1,k}}$  such that (i)  $w$  belongs to the right-hand side of (1.4), (ii)  $w'$  does not belong to the right-hand side of (1.4) and (iii)  $w \langle \cdot \rangle_v^{1,k} w'$ . This means  $w = xvz$  and  $w' = xvuvz$  for suitable  $x, z \in A^*$  and  $u \in A^{*\geq k+1}$  with  $\alpha_k(vuv) \subseteq \alpha_k(v)$ . On the other hand, since  $w$  is an element of the right-hand side of (1.4), there exist suitable  $x_2, z_2, \beta', \delta'$  such that  $w \in \langle x_1 x_2 \rangle_{\langle \cdot \rangle_v^{1,k}} \cdot (\beta'|\Gamma|\delta')_k \cdot \langle z_2 z_1 \rangle_{\langle \cdot \rangle_v^{1,k}}$ . So we have  $w = \tilde{x}\tilde{v}\tilde{z}$  for suitable  $\tilde{x} \in \langle x_1 x_2 \rangle_{\langle \cdot \rangle_v^{1,k}}$ ,  $\tilde{v} \in (\beta'|\Gamma|\delta')_k$  and  $\tilde{z} \in \langle z_2 z_1 \rangle_{\langle \cdot \rangle_v^{1,k}}$ . Now we compare the decompositions  $w = xvz$  and  $w = \tilde{x}\tilde{v}\tilde{z}$ .

**Case 1:** The factor  $v$  of the decomposition  $w = xvz$  is contained in  $\tilde{x}$  or in  $\tilde{z}$  of the decomposition  $w = \tilde{x}\tilde{v}\tilde{z}$ . Without loss of generality we assume that  $v$  is a factor of  $\tilde{x}$ .



Note that  $\tilde{x} \langle \cdot \rangle_v^{1,k} \tilde{x}'$  where  $\tilde{x}'$  is defined as in the picture above. It follows that  $\tilde{x}' \in \langle x_1 x_2 \rangle_{\langle \cdot \rangle_v^{1,k}}$ , and therefore  $w'$  is an element of the right-hand side of (1.4). This is a contradiction.

**Case 2:** The factor  $v$  of the decomposition  $w = x v z$  is neither contained in  $\tilde{x}$  nor contained in  $\tilde{z}$ . Hence,  $v$  is contained in  $\mathfrak{s}_{|v|}(\tilde{x}) \tilde{v} \mathfrak{p}_{|v|}(\tilde{z})$ . By Proposition 1.20 it holds that  $\mathfrak{s}_{|v|}(\tilde{x}) = x_2$  and  $\mathfrak{p}_{|v|}(\tilde{z}) = z_2$ . This yields the following decompositions of  $w$  and  $w'$ .



Clearly, it holds that  $y \langle \cdot \rangle_v^{1,k} y'$  where  $y$  and  $y'$  are defined as in the picture. By assumption we have  $\alpha_k(x_2 \beta') \subseteq \Gamma$ ,  $\alpha_k(\delta' z_2) \subseteq \Gamma$  and  $\tilde{v} \in (\beta' | \Gamma | \delta')_k$ . This shows  $|y| \geq 2|v| + k + 1$  and  $\alpha_k(y) \subseteq \Gamma$ . Hence, from Proposition 1.20 it follows that  $\alpha_k(y') = \alpha_k(y) \subseteq \Gamma$ . In particular there exist  $\beta'', \delta'' \in A^k$  with  $\alpha_k(x_2 \beta''), \alpha_k(\delta'' z_2) \subseteq \Gamma$  such that  $y' \in x_2 (\beta'' | \Gamma | \delta'')_k z_2$ . This means that  $w' \in \tilde{x} (\beta'' | \Gamma | \delta'')_k \tilde{z}$ , and it follows that  $w' \in \langle x_1 x_2 \rangle_{\langle \cdot \rangle_v^{1,k}} \cdot (\beta'' | \Gamma | \delta'')_k \cdot \langle z_2 z_1 \rangle_{\langle \cdot \rangle_v^{1,k}}$ . This is a contradiction since we assumed that  $w'$  does not belong to the right-hand side of (1.4).

So in all possible cases we obtain contradictions. Hence, our assumption was false and equation (1.4) follows. Together with Corollary 1.27 we get  $\langle L \cap A^{*\geq n} \rangle_{\langle \cdot \rangle_v^{1,k}} \in \tilde{\mathcal{B}}_k$ .  $\square$

#### 1.4.5 The $\langle \cdot \rangle_v^{1,k}$ Upward Closure of Languages from $\tilde{\mathcal{B}}_k$

With Lemma 1.28 we prepared the induction base for Theorem 1.30—the main result of this section. The corresponding induction step is prepared with the following decomposition lemma.

**Lemma 1.29.** *Let  $k \geq 0$ ,  $v \in A^{*\geq k+1}$  and  $L_1, L_2 \subseteq A^*$ . Then it holds that*

$$\langle L_1 L_2 \rangle_{\langle \cdot \rangle_v^{1,k}} = \langle L_1 \rangle_{\langle \cdot \rangle_v^{1,k}} \cdot \langle L_2 \rangle_{\langle \cdot \rangle_v^{1,k}} \cup \bigcup_{\substack{v_1, v_2 \in A^*, \\ v = v_1 v_2}} \langle (L_1 \cap A^* v_1) v_2 \rangle_{\langle \cdot \rangle_v^{1,k}} \cdot L_{k,v} \cdot \langle v_1 (L_2 \cap v_2 A^*) \rangle_{\langle \cdot \rangle_v^{1,k}}. \quad (1.5)$$

*Proof.* We start with the proof showing that the right-hand side is contained in the left-hand side. By Proposition 1.20.6 we have  $\langle L_1 L_2 \rangle_{\langle \cdot \rangle_v^{1,k}} \supseteq \langle L_1 \rangle_{\langle \cdot \rangle_v^{1,k}} \cdot \langle L_2 \rangle_{\langle \cdot \rangle_v^{1,k}}$ . So it remains to show that  $\langle L_1 L_2 \rangle_{\langle \cdot \rangle_v^{1,k}} \supseteq \langle (L_1 \cap A^* v_1) v_2 \rangle_{\langle \cdot \rangle_v^{1,k}} \cdot L_{k,v} \cdot \langle v_1 (L_2 \cap v_2 A^*) \rangle_{\langle \cdot \rangle_v^{1,k}}$  for  $v_1, v_2 \in A^*$  with  $v = v_1 v_2$ . Since  $\langle L_1 L_2 \rangle_{\langle \cdot \rangle_v^{1,k}}$  is closed under  $\langle \cdot \rangle_v^{1,k}$  it suffices to show

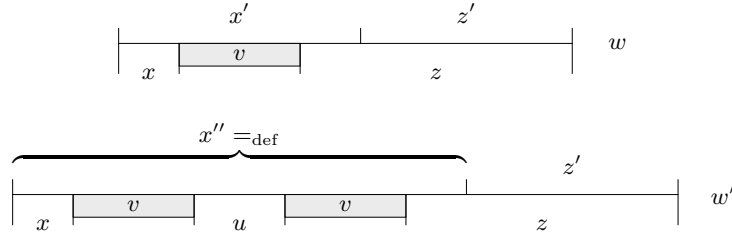
$$\langle L_1 L_2 \rangle_{\langle \cdot \rangle_v^{1,k}} \supseteq (L_1 \cap A^* v_1) v_2 \cdot L_{k,v} \cdot v_1 (L_2 \cap v_2 A^*). \quad (1.6)$$

If  $w' \in (L_1 \cap A^*v_1)v_2 \cdot L_{k,v} \cdot v_1(L_2 \cap v_2A^*)$  then there exist  $u \in L_{k,v}$ ,  $x, z \in A^*$  with  $xv_1 \in L_1$  and  $v_2z \in L_2$  such that  $w' = xv_1v_2uv_1v_2z$ . It follows that  $w =_{\text{def}} xv_1v_2z \in L_1L_2$ . Since  $w = xvz$ ,  $w' = xvvuz$  and  $u \in L_{k,v}$ , we obtain  $w <_v^{1,k} w'$ . This shows (1.6).

We turn to the inclusion from left to right in equation (1.5). Assume that this inclusion does not hold, this will lead to a contradiction. We choose a word  $w'$  of minimal length such that  $w' \in \langle L_1L_2 \rangle_{<_v^{1,k}}$  and  $w'$  does not belong to the right-hand side of (1.5). Since  $L_1L_2 \subseteq \langle L_1 \rangle_{<_v^{1,k}} \cdot \langle L_2 \rangle_{<_v^{1,k}}$ , we have  $w' \notin L_1L_2$ . Hence there exists a word  $w \in \langle L_1L_2 \rangle_{<_v^{1,k}}$  with  $w <_v^{1,k} w'$ . From the minimal choice of  $w'$  it follows that  $w$  is an element of the right-hand side of (1.5). So there exist words  $x, z \in A^*$  and  $u \in L_{k,v}$  such that  $w = xvz$  and  $w' = xvvuz$ .

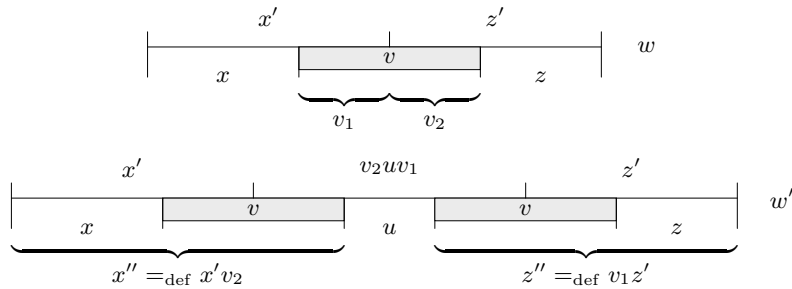
**Case 1:** Assume that  $w \in \langle L_1 \rangle_{<_v^{1,k}} \cdot \langle L_2 \rangle_{<_v^{1,k}}$ . Let  $x' \in \langle L_1 \rangle_{<_v^{1,k}}$  and  $z' \in \langle L_2 \rangle_{<_v^{1,k}}$  such that  $w = x'z'$ .

*Case 1a:* Assume that the factor  $v$  of the decomposition  $w = xvz$  is either contained in the factor  $x'$  or is contained in the factor  $z'$  of the decomposition  $w = x'z'$ . Without loss of generality we assume here that  $v$  is contained in the factor  $x'$ .



Let  $x''$  as in the picture. We obtain  $x' <_v^{1,k} x''$ , and it follows that  $w' \in \langle L_1 \rangle_{<_v^{1,k}} \cdot \langle L_2 \rangle_{<_v^{1,k}}$ . This contradicts our assumption.

*Case 1b:* Assume that the factor  $v$  of the decomposition  $w = xvz$  overlaps both factors  $x'$  and  $z'$  of the decomposition  $w = x'z'$ . Then there exists a decomposition  $v = v_1v_2$  such that  $v_1$  is the prefix of  $v$  that overlaps  $x'$ , and  $v_2$  is the suffix of  $v$  that overlaps  $z'$ . So we have the following situation.



Since  $x' \in \langle L_1 \rangle_{<_v^{1,k}}$  there exists a  $y' \in L_1$  with  $x' \in \langle y' \rangle_{<_v^{1,k}}$ . From Proposition 1.20 it follows that  $y'$  has the suffix  $v_1$  and that  $x'v_2 \in \langle y'v_2 \rangle_{<_v^{1,k}}$ . Hence  $x'' = x'v_2 \in \langle (L_1 \cap A^*v_1)v_2 \rangle_{<_v^{1,k}}$ ,

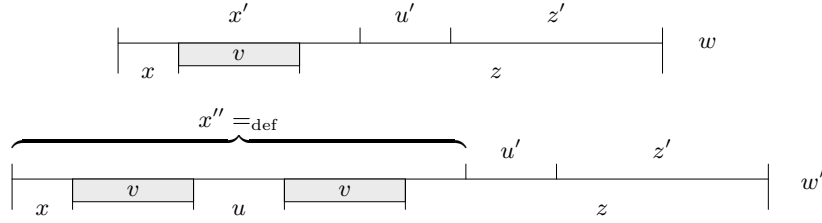
and analogously we obtain  $z'' \in \langle v_1(L_2 \cap v_2A^*) \rangle_{\prec_v^{1,k}}$ . From  $u \in L_{k,v}$  it follows that

$$w' = x''uz'' \in \langle (L_1 \cap A^*v_1)v_2 \rangle_{\prec_v^{1,k}} \cdot L_{k,v} \cdot \langle v_1(L_2 \cap v_2A^*) \rangle_{\prec_v^{1,k}}.$$

This contradicts the assumption that  $w'$  does not belong to the right-hand side of (1.5).

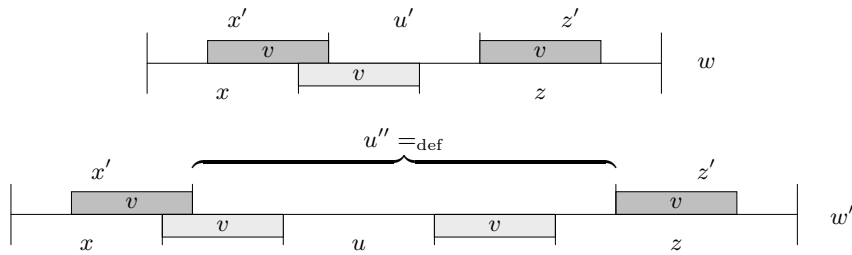
**Case 2:** Assume that  $w \in \langle (L_1 \cap A^*v_1)v_2 \rangle_{\prec_v^{1,k}} \cdot L_{k,v} \cdot \langle v_1(L_2 \cap v_2A^*) \rangle_{\prec_v^{1,k}}$  for suitable  $v_1, v_2 \in A^*$  with  $v = v_1v_2$ . Let  $x' \in \langle (L_1 \cap A^*v_1)v_2 \rangle_{\prec_v^{1,k}}$ ,  $z' \in \langle v_1(L_2 \cap v_2A^*) \rangle_{\prec_v^{1,k}}$  and  $u' \in L_{k,v}$  such that  $w = x'u'z'$ .

*Case 2a:* Assume that the factor  $v$  of the decomposition  $w = xvz$  is either contained in the factor  $x'$  or is contained in the factor  $z'$  of the decomposition  $w = x'u'z'$ . Without loss of generality we assume here that  $v$  is contained in the factor  $x'$ .



Define  $x''$  as in the picture, and observe that  $x' \prec_v^{1,k} x''$ . Hence  $x'' \in \langle (L_1 \cap A^*v_1)v_2 \rangle_{\prec_v^{1,k}}$ , and it follows that  $w' \in \langle (L_1 \cap A^*v_1)v_2 \rangle_{\prec_v^{1,k}} \cdot L_{k,v} \cdot \langle v_1(L_2 \cap v_2A^*) \rangle_{\prec_v^{1,k}}$ . This is a contradiction to our assumption.

*Case 2b:* Assume that the factor  $v$  of the decomposition  $w = xvz$  is contained in the factor  $\mathfrak{s}_{|v|}(x')u'\mathfrak{p}_{|v|}(z')$  of the decomposition  $w = x'u'z'$ . Since  $x' \in \langle (L_1 \cap A^*v_1)v_2 \rangle_{\prec_v^{1,k}}$  there exists a  $y' \in (L_1 \cap A^*v_1)v_2$  such that  $x' \in \langle y' \rangle_{\prec_v^{1,k}}$ . Note that  $y'$  has the suffix  $v$ . From Proposition 1.20 it follows that also  $x'$  has the suffix  $v$ , and analogously we get that  $z'$  has the prefix  $v$ . So the following situation emerges.



Since  $u' \in L_{k,v}$  we have  $\alpha_k(vu'v) = \alpha_k(v)$ . Moreover, we see from the picture above that  $vu'v \prec_v^{1,k} vu''v$ . From Proposition 1.20 it follows that  $\alpha_k(vu''v) = \alpha_k(vu'v) = \alpha_k(v)$ . So we obtain  $u'' \in L_{k,v}$  and  $w' \in \langle (L_1 \cap A^*v_1)v_2 \rangle_{\prec_v^{1,k}} \cdot L_{k,v} \cdot \langle v_1(L_2 \cap v_2A^*) \rangle_{\prec_v^{1,k}}$ . This contradicts our assumption that  $w'$  does not belong to the right-hand side of (1.5).

We have seen that in all possible cases we get contradictions. So our assumption was false, and it follows that in (1.5) the left-hand side is a subset of the right-hand side.  $\square$

Finally, we state the main theorem of this section. It says that the  $\langle \cdot \rangle_{\prec_v^{1,k}}$  upward closure of a language from  $\tilde{\mathcal{B}}_k$  is in  $\tilde{\mathcal{B}}_k$ . As a corollary we obtain the following: If we start from a single word  $w$  and if we take several  $\langle \cdot \rangle_{\prec_v^{1,k}}$  upward closures then we are still in  $\tilde{\mathcal{B}}_k$  (and therefore, as we will see in chapter 4, we are in the level 3/2 of the dot-depth hierarchy).

**Theorem 1.30.** *Let  $k \geq 0$ ,  $v \in A^{*\geq k+1}$  and  $L \in \tilde{\mathcal{B}}_k$ . Then it holds that  $\langle L \rangle_{\prec_v^{1,k}} \in \tilde{\mathcal{B}}_k$ .*

*Proof.* Since the classes  $\tilde{\mathcal{B}}_{k,m}$  exhaust  $\tilde{\mathcal{B}}_k$  it suffices to show the following claim.

**Claim.** *Let  $m \geq 1$  and  $L \in \tilde{\mathcal{B}}_{k,m}$ . Then it holds that  $\langle L \rangle_{\prec_v^{1,k}} \in \tilde{\mathcal{B}}_k$ .*

We prove the claim by induction on  $m$ . The induction base (i.e., the case  $m = 1$ ) follows from Lemma 1.28. So we assume that the lemma has been proved for  $m = l \geq 1$ , and we want to show it for  $m = l + 1$ .

By definition, languages from  $\tilde{\mathcal{B}}_{k,m}$  are finite unions of languages  $L = L_1 L_2 \cdots L_n$  with  $1 \leq n \leq l + 1$  such that for  $1 \leq i \leq n$  it holds that either  $L_i = \{w_i\}$  for some  $w_i \in A^+$ , or  $L_i = x_i(\beta_i|\Gamma_i|\delta_i)_k z_i$  for  $x_i, z_i \in A^*$ ,  $\beta_i, \delta_i \in A^k$  and  $\Gamma_i \subseteq A^{k+1}$ . Therefore, it suffices to show  $\langle L \rangle_{\prec_v^{1,k}} \in \tilde{\mathcal{B}}_k$ . For  $n \leq l$  this follows from the induction hypothesis. So we assume that  $n = l + 1$ . Let  $L' =_{\text{def}} L_1 L_2 \cdots L_l$  and observe that  $L', L_{l+1} \in \tilde{\mathcal{B}}_{k,l}$ . By Lemma 1.29 we have

$$\langle L \rangle_{\prec_v^{1,k}} = \langle L' \rangle_{\prec_v^{1,k}} \cdot \langle L_{l+1} \rangle_{\prec_v^{1,k}} \cup \bigcup_{\substack{v_1, v_2 \in A^*, \\ v = v_1 v_2}} \langle (L' \cap A^* v_1) v_2 \rangle_{\prec_v^{1,k}} \cdot L_{k,v} \cdot \langle v_1 (L_{l+1} \cap v_2 A^*) \rangle_{\prec_v^{1,k}}.$$

From the induction hypothesis it follows that  $\langle L' \rangle_{\prec_v^{1,k}} \cdot \langle L_{l+1} \rangle_{\prec_v^{1,k}} \in \tilde{\mathcal{B}}_k$ . Let  $v_1, v_2 \in A^*$  with  $v = v_1 v_2$ . By Proposition 1.25 we have  $(L' \cap A^* v_1) v_2 \in \tilde{\mathcal{B}}_{k,l}$  and  $v_1 (L_{l+1} \cap v_2 A^*) \in \tilde{\mathcal{B}}_{k,l}$ . From the induction hypothesis we obtain  $\langle (L' \cap A^* v_1) v_2 \rangle_{\prec_v^{1,k}} \in \tilde{\mathcal{B}}_k$  and  $\langle v_1 (L_{l+1} \cap v_2 A^*) \rangle_{\prec_v^{1,k}} \in \tilde{\mathcal{B}}_k$ . With  $L_{k,v} \in \tilde{\mathcal{B}}_{k,1} \subseteq \tilde{\mathcal{B}}_k$  this implies  $\langle (L' \cap A^* v_1) v_2 \rangle_{\prec_v^{1,k}} \cdot L_{k,v} \cdot \langle v_1 (L_{l+1} \cap v_2 A^*) \rangle_{\prec_v^{1,k}} \in \tilde{\mathcal{B}}_k$ , and we conclude that  $\langle L \rangle_{\prec_v^{1,k}} \in \tilde{\mathcal{B}}_k$ .  $\square$

**Corollary 1.31.** *Let  $k, n \geq 0$ ,  $v_0, \dots, v_n \in A^{*\geq k+1}$  and  $w \in A^+$ . Then it holds that*

$$\langle \cdots \langle \langle w \rangle_{\prec_{v_0}^{1,k}} \rangle_{\prec_{v_1}^{1,k}} \cdots \rangle_{\prec_{v_n}^{1,k}} \in \tilde{\mathcal{B}}_k.$$

*Proof.* This follows from Theorem 1.30 since  $\{w\} \in \tilde{\mathcal{B}}_k$ .  $\square$



## 2. Concatenation Hierarchies

This chapter introduces the notion of concatenation hierarchies and in particular that of the famous dot-depth hierarchy. For this we start in section 2.1 with the definition of the polynomial closure of a class of languages. Then we define the dot-depth hierarchy and the Straubing-Thérien hierarchy which are well-known concatenation hierarchies. We prove easy inclusion relations and state that both hierarchies exhaust the class of starfree languages.

In section 2.2 we discuss alternative definitions for the considered concatenation hierarchies. It will turn out that the use of alternative definitions lead to minor changes which can be neglected. In particular, the essential decidability questions concerning these hierarchies remain the same.

In section 2.3 we give some useful normalforms for concatenation hierarchies, and in section 2.4 the famous dot-depth problem is stated.

Finally, section 2.5 introduces the notion of forbidden-pattern characterizations and summarizes known characterizations for the levels  $n + 1/2$  of concatenation hierarchies. We compare the corresponding patterns to get an idea of their general structure.

### 2.1 Definitions of Concatenation Hierarchies

Regular languages are constructed from alphabet letters by the use of Boolean operations (i.e., finite union, finite intersection and complementation), concatenation and iteration. Ignoring iterations, the class of *starfree languages* (SF for short) is defined as the smallest class of languages that contains the atomic languages  $\{a\}$  for  $a \in A$  and that is closed under finite Boolean operations and concatenation.

A systematic way to examine this class is to count the number of alternating uses of Boolean operations on the one hand and concatenations on the other hand. This means that for a starfree language we consider the number of unavoidable alternations between combinatorial and sequential aspects in the definition of the language. For a given language we call this number the *concatenation complexity*.

Grouping together languages of similar concatenation complexity leads in a natural way to the definition of *concatenation hierarchies* that exhaust the class of starfree languages. Prominent examples are the dot-depth hierarchy (DDH), first studied in [CB71], and the Straubing-Thérien hierarchy (STH) [Str81, Thé81, Str85]. In contrast to the original definitions for the DDH and STH, in this thesis we use alternative (slightly modified) versions. See section 2.2 for a discussion of this.

For a class  $\mathcal{C}$  of languages we denote its closure under finite (possibly empty) union by  $\text{FU}(\mathcal{C})$ . The polynomial closure of  $\mathcal{C}$  is defined as

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

Note that  $\text{Pol}(\mathcal{C})$  is exactly the closure of  $\mathcal{C}$  under finite (possibly empty) union and finite (nonempty) concatenation. Observe that  $\mathcal{C}$  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  $\text{BC}(\mathcal{C})$  (this means that we take complements with respect to  $A^+$ ).

**Definition 2.1 (DDH).** *The classes of the dot-depth hierarchy are defined as*

$$\begin{aligned} \mathcal{B}_{1/2} &=_{\text{def}} \text{Pol}(\{\{a\} \mid a \in A\} \cup \{A^+\}), \\ \mathcal{B}_{n+1} &=_{\text{def}} \text{BC}(\mathcal{B}_{n+1/2}) && \text{for } n \geq 0 \text{ and} \\ \mathcal{B}_{n+3/2} &=_{\text{def}} \text{Pol}(\mathcal{B}_{n+1}) && \text{for } n \geq 0. \end{aligned}$$

**Definition 2.2 (STH).** *The classes of the Straubing-Thérien hierarchy are defined as*

$$\begin{aligned} \mathcal{L}_{1/2} &=_{\text{def}} \text{Pol}(\{A^* a A^* \mid a \in A\}), \\ \mathcal{L}_{n+1} &=_{\text{def}} \text{BC}(\mathcal{L}_{n+1/2}) && \text{for } n \geq 0 \text{ and} \\ \mathcal{L}_{n+3/2} &=_{\text{def}} \text{Pol}(\mathcal{L}_{n+1}) && \text{for } n \geq 0. \end{aligned}$$

Note that this defines indeed classes of starfree languages since  $A^+$  is the complement of the empty set, and since  $A^* a A^*$  is equal to  $\{a\} \cup A^+ a \cup a A^+ \cup A^+ a A^+$ . We call the introduced classes also the levels of the DDH and STH where  $\mathcal{B}_n, \mathcal{L}_n$  are the full levels and  $\mathcal{B}_{n-1/2}, \mathcal{L}_{n-1/2}$  are the half levels for integers  $n \geq 1$ . From the definitions above we immediately obtain the following inclusion structure.

**Proposition 2.3.** *For  $n \geq 0$  the following holds.*

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

Moreover, we have also inclusion relations between both hierarchies.

**Proposition 2.4.** *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.  $\text{co}\mathcal{L}_{n-1/2} \subseteq \text{co}\mathcal{B}_{n-1/2} \subseteq \text{co}\mathcal{L}_{n+1/2}$
3.  $\mathcal{L}_n \subseteq \mathcal{B}_n \subseteq \mathcal{L}_{n+1}$

*Proof.* Since  $A^* a A^* = \{a\} \cup A^+ a \cup a A^+ \cup A^+ a A^+$  for all  $a \in A$  we get  $\mathcal{L}_{1/2} \subseteq \mathcal{B}_{1/2}$ . Note that  $A^+ = \bigcup_{a \in A} A^* a A^* \in \mathcal{L}_{1/2} \subseteq \mathcal{L}_{3/2}$ . Moreover, for  $w \in A^+$  with  $w = a_1 \cdots a_n$  for  $n \geq 1$  and letters  $a_i \in A$  we obtain

$$\{w\} = \underbrace{A^* a_1 A^* \cdots a_n A^*}_{\in \mathcal{L}_{1/2}} \cap \left( A^+ \setminus \underbrace{\bigcup_{b_1, \dots, b_{n+1} \in A} A^* b_1 A^* \cdots b_{n+1} A^*}_{\in \mathcal{L}_{1/2}} \right) \in \mathcal{L}_1 \subseteq \mathcal{L}_{3/2}.$$

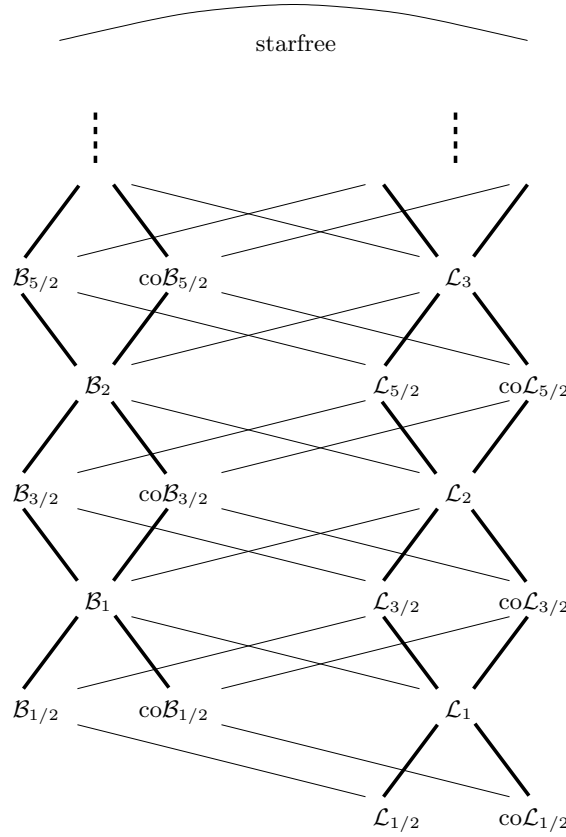


Fig. 2.1. Inclusion Relations of the DDH and STH

It follows that  $\mathcal{B}_{1/2} \subseteq \mathcal{L}_{3/2}$ . 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  $\text{Pol}(\cdot)$ ,  $\text{BC}(\cdot)$ , and complementation.  $\square$

The classes  $\mathcal{B}_n$  for  $n \geq 1$  coincide with the ones studied in [Eil76]. In [Eil76, chapter IX.4] it is shown that  $\bigcup_{n \geq 1} \mathcal{B}_n = \text{SF}$ . Together with Proposition 2.4 this shows the following.

**Theorem 2.5 ([Eil76]).**  $\bigcup_{n \geq 1} \mathcal{L}_{n/2} = \bigcup_{n \geq 1} \mathcal{B}_{n/2} = \text{SF}$

Figure 2.1 shows the inclusion relations that we obtained so far. Beside the inclusion relations there are also strictness results. In [BK78, Tho84, Str85] it is shown that the DDH and the STH are strict. This implies the strictness of all inclusions shown in Figure 2.1. Moreover, several characterizations for these hierarchies are known, e.g., in the theory of finite semigroups, in finite model theory, and in complexity theory. For an overview of this rich field of current interest and research see, e.g., [Brz76, Pin96a, Pin96b].

## 2.2 Alternative Definitions

The DDH and the STH have gained much attention due to the still pending dot-depth problem (see Problem 2.12). The purpose of this section is to make our work comparable to other investigations, i.e., it relates our alternative definitions of the DDH and STH to the ones known from literature. We will see that in the case of the DDH our definition describes exactly the original classes. For the STH these classes (except level 1/2) differ in exactly the empty word.

Essentially, there are three differences between the original definitions and our alternative ones. First, the STH is defined a way such that their languages may contain the empty word. A second point is that one uses other versions of the polynomial closure operation. Finally, the original definitions of both hierarchies start with level 0 (and not with level 1/2) which is defined as the class that contains the empty language and the language of all words (either  $A^+$  or  $A^*$ ). So the latter difference is a minor one and we will neglect it in the further considerations. In particular we will give the original definitions only for levels greater than or equal to 1/2.

Let  $\mathcal{C}$  be a class of languages. In the literature, the DDH contains languages from  $A^+$  and the STH contains languages from  $A^*$ . Because of this difference one has to use different versions of the polynomial closure operation.

$$\begin{aligned} \text{Pol}^{\mathcal{L}}(\mathcal{C}) &=_{\text{def}} \text{FU}(\{L_0 a_1 L_1 \cdots a_n L_n : n \geq 0, L_i \in \mathcal{C} \text{ and } a_i \in A\}) \\ \text{Pol}^{\mathcal{B}}(\mathcal{C}) &=_{\text{def}} \text{FU}(\{u_0 L_1 u_1 \cdots L_n u_n : n \geq 0, L_i \in \mathcal{C}, u_i \in A^* \text{ and if } n = 0 \text{ then } u_0 \neq \varepsilon\}) \end{aligned}$$

It is pointed out, e.g., in [Pin95] that this is a crucial point in the theory of varieties of finite semigroups. Since many 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 from  $A^*$  by  $\text{BC}^*(\mathcal{D})$  (i.e., we take complements with respect to  $A^*$ ). Moreover, let  $\text{co}^*\mathcal{D} =_{\text{def}} \{A^* \setminus L \mid L \in \mathcal{D}\}$  denote the set of complements with respect to  $A^*$ .

**Definition 2.6 (DDH due to [Pin96b]).** 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  $u_0 A^+ u_1 \cdots A^+ u_m$  where  $m \geq 0$  and  $u_i \in A^*$ . For  $n \geq 0$  let  $\mathcal{B}_{n+1}^+ =_{\text{def}} \text{BC}^*(\mathcal{B}_{n+1/2}^+)$  and  $\mathcal{B}_{n+3/2}^+ =_{\text{def}} \text{Pol}^{\mathcal{B}}(\mathcal{B}_{n+1}^+)$ .

Also the definition above differs a bit from those given in earlier literature. The levels  $n + 1/2$  extend the corresponding classes defined in [CB71], however the levels  $n$  coincide.

**Definition 2.7 (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  $A^* a_1 A^* \cdots a_m A^*$  where  $m \geq 0$  and  $a_i \in A$ . For  $n \geq 0$  let  $\mathcal{L}_{n+1}^* =_{\text{def}} \text{BC}^*(\mathcal{L}_{n+1/2}^*)$  and  $\mathcal{L}_{n+3/2}^* =_{\text{def}} \text{Pol}^{\mathcal{L}}(\mathcal{L}_{n+1}^*)$ .

The theorem below shows that the classes  $\mathcal{B}_{n/2}^+$  from Definition 2.6 and our classes  $\mathcal{B}_{n/2}$  coincide, and that for  $n \geq 2$  the languages from  $\mathcal{L}_{n/2}^*$  are up to the empty word the languages from  $\mathcal{L}_{n/2}$ .

**Theorem 2.8** ([GS00b]). *The following holds for  $n \geq 1$  and  $m \geq 2$ .*

1.  $\mathcal{B}_{n/2}^+ = \mathcal{B}_{n/2}$
2.  $\text{co}\mathcal{B}_{n/2}^+ = \text{co}\mathcal{B}_{n/2}$
3.  $\mathcal{L}_{1/2}^* = \mathcal{L}_{1/2} \cup \{A^*\}$
4.  $\mathcal{L}_{m/2}^* = \mathcal{L}_{m/2} \cup \{L \cup \{\varepsilon\} \mid L \in \mathcal{L}_{m/2}\}$
5.  $\text{co}^*\mathcal{L}_{m/2}^* = \text{co}\mathcal{L}_{m/2} \cup \{L \cup \{\varepsilon\} \mid L \in \text{co}\mathcal{L}_{m+1/2}\}$

## 2.3 Normalforms and Closure Properties

With help of Theorem 2.8 we can take over existing normalforms and closure properties for the DDH and STH.

**Theorem 2.9** ([Arf91]).  $\mathcal{L}_{3/2} = \text{Pol}(\{B^+ \mid B \subseteq A\} \cup \{\{a\} \mid a \in A\})$ .

**Theorem 2.10** ([Gla98]). *For  $n \geq 1$  the following holds.*

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

Finally, with Theorem 2.8 we can translate the following facts from [Arf91, PW97].

**Theorem 2.11.** *Let  $n \geq 1$  and  $a \in A$ .*

1.  $\mathcal{B}_{n/2}, \text{co}\mathcal{B}_{n/2}, \mathcal{L}_{n/2}$  and  $\text{co}\mathcal{L}_{n/2}$  are closed under finite union and finite intersection.
2. If  $\mathcal{C}$  is one of the classes  $\mathcal{B}_{n/2}, \text{co}\mathcal{B}_{n/2}, \mathcal{L}_{n/2}$  or  $\text{co}\mathcal{L}_{n/2}$  then for all  $L \in \mathcal{C}$  it holds that  $a^{-1}L \cap A^+ \in \mathcal{C}$  and  $La^{-1} \cap A^+ \in \mathcal{C}$ .

## 2.4 The Dot-Depth Problem

The main motivation for dealing with concatenation hierarchies like the DDH and STH comes from an easy explainable problem.

**Problem 2.12 (dot-depth problem).** *Does there exist an algorithm that computes on input of a given starfree language  $L \in A^+$  the minimal  $n \geq 1$  such that  $L \in \mathcal{B}_{n/2}$ ?*

A lot of effort in different approaches has been invested to solve this problem. However, 30 years after its discovery, the dot-depth problem is still open. Not least owing to this, it is sometimes called the P-NP problem of the automata theory. It remains an extremely difficult problem even if we restrict it in such a way that we ask for an algorithm that decides the membership problem for a single class (e.g.,  $\mathcal{B}_2$ ).

In section 2.1 we have seen structural similarities and inclusion relations of the DDH and the STH. The following theorem shows that both hierarchies are also similar when looking at the decidability of their membership problems.

**Theorem 2.13** ([Str85]). *For every  $n \geq 1$ ,  $\mathcal{L}_n$  is decidable if and only if  $\mathcal{B}_n$  is decidable.*

Recently, it was shown in [PW01] that this theorem even holds for all levels  $n + 1/2$ .

**Theorem 2.14 ([PW01]).** *For every  $n \geq 0$ ,  $\mathcal{L}_{n+1/2}$  is decidable if and only if  $\mathcal{B}_{n+1/2}$  is decidable.*

Until recently, only the levels 1/2 and 1 of the DDH [Kna83, PW97], and the levels 1/2, 1 and 3/2 of the STH [Sim75, Arf91, PW97] were known to be decidable.

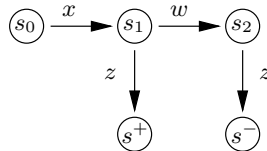
The decidability of level 3/2 of the DDH can now be obtained in three ways. First, by an automata-theoretic approach using forbidden-patterns [GS00a]. At second, using the decidability of level 3/2 of the STH together with Theorem 2.14 from [PW01]. In this thesis we offer a third approach which is also of automata-theoretic nature but which uses word extensions.

For a detailed overview on the strong influence of the long-standing open decidability questions and on the continuously ongoing research see [Pin95, Pin96b].

## 2.5 Known Forbidden-Pattern Characterizations

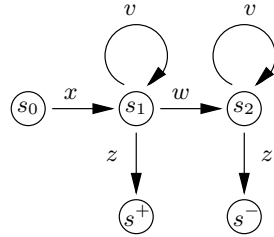
Most of the known decidability results for the classes of the DDH and STH are of the following type: “ $L$  belongs to the class if and only if the accepting DFA does not have subgraph  $\mathbb{P}$  in its transition graph”. Results of this type are called *forbidden-pattern characterizations*, and usually they imply the decidability of the characterized classes (see chapter 3). To our knowledge such forbidden-pattern characterizations are known for  $\mathcal{L}_{1/2}$ ,  $\mathcal{L}_1$ ,  $\mathcal{L}_{3/2}$ ,  $\mathcal{B}_{1/2}$  and  $\mathcal{B}_1$  [Sim75, Kna83, Arf91, PW97], and in chapter 4 we will prove one for  $\mathcal{B}_{3/2}$  [GS00a].

In this thesis we restrict our attention to characterizations of the levels  $n+1/2$  (these are the levels that are closed under concatenation). The Figures 2.2–2.4 show the forbidden-pattern characterizations for  $\mathcal{L}_{1/2}$ ,  $\mathcal{B}_{1/2}$  and  $\mathcal{L}_{3/2}$  [Arf91, PW97]. In Figure 2.5 we anticipate the characterization for  $\mathcal{B}_{3/2}$  which we will prove in chapter 4.

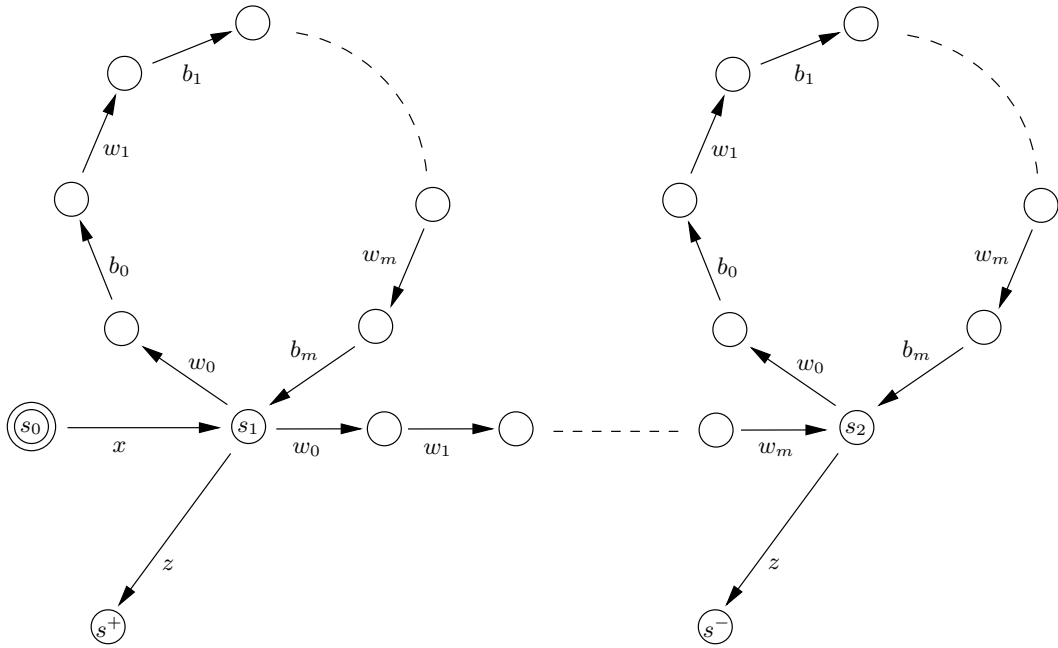


**Fig. 2.2.** Forbidden-pattern for  $\mathcal{L}_{1/2}$  [Arf91, PW97] with initial state  $s_0$ , accepting state  $s^+$ , rejecting state  $s^-$  and words  $x, w, z \in A^*$ .

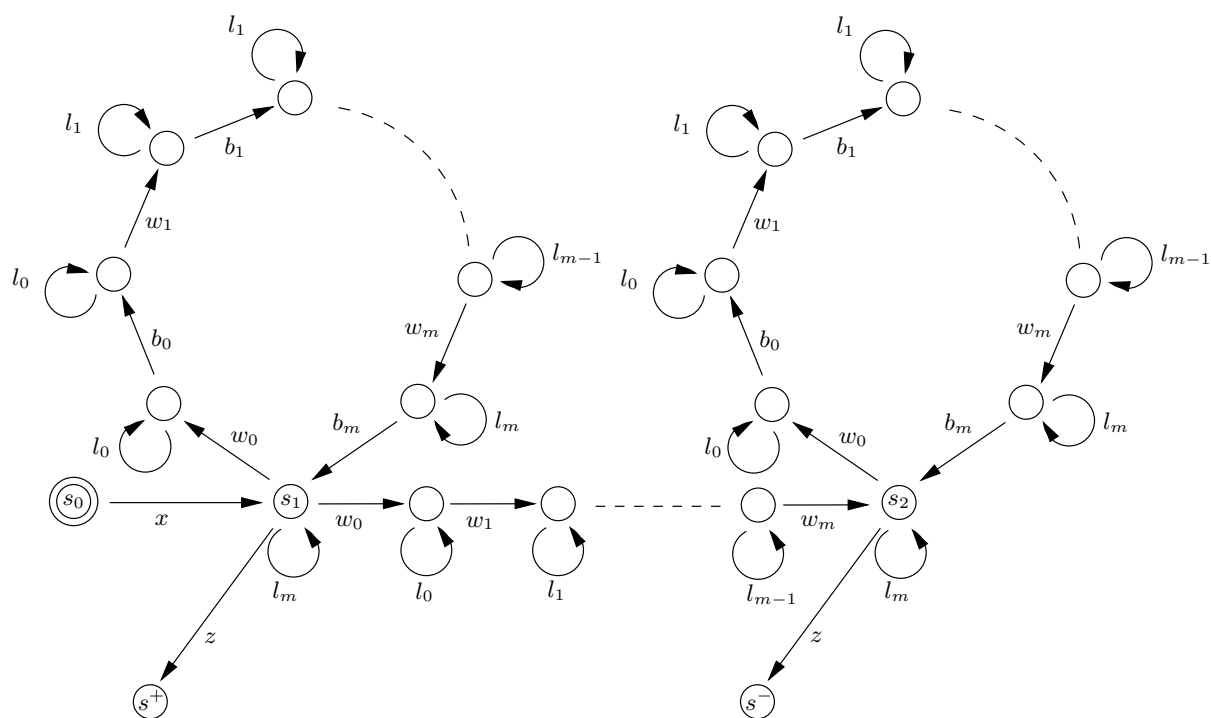
If we compare the patterns in the Figures 2.2–2.5 then we observe that they are of the following form (cf. Figure 2.6): There appear states  $s_1, s_2$  and words  $x, z \in A^*$  such that (i)  $x$  leads from the initial state to  $s_1$ , (ii)  $z$  leads from  $s_1$  to an accepting state  $s^+$  and from  $s_2$  to a rejecting state  $s^-$  and (iii) we find a certain structure between  $s_1$  and  $s_2$  (the gray area in Figure 2.6). Hence in order to define a certain pattern it suffices to describe the structure between  $s_1$  and  $s_2$ , i.e., the gray area in the picture. Therefore, from now on we will neglect the words  $x, z$  and the states  $s_0, s^+, s^-$ .



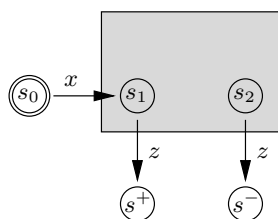
**Fig. 2.3.** Forbidden-pattern for  $\mathcal{B}_{1/2}$  [PW97] with initial state  $s_0$ , accepting state  $s^+$ , rejecting state  $s^-$  and words  $x, z \in A^*$ ,  $v, w \in A^+$ .



**Fig. 2.4.** Forbidden-pattern for  $\mathcal{L}_{3/2}$  [PW97] with  $m \geq 0$ , initial state  $s_0$ , accepting state  $s^+$ , rejecting state  $s^-$  and words  $x, z, b_i \in A^*$ ,  $w_i \in A^+$ .



**Fig. 2.5.** Forbidden-pattern for  $\mathcal{B}_{3/2}$  [GS00a] with  $m \geq 0$ , initial state  $s_0$ , accepting state  $s^+$ , rejecting state  $s^-$  and words  $x, z \in A^*$ ,  $b_i, l_i, w_i \in A^+$ .



**Fig. 2.6.** General structure of the forbidden-patterns for  $\mathcal{L}_{1/2}$ ,  $\mathcal{L}_{3/2}$ ,  $\mathcal{B}_{1/2}$  and  $\mathcal{B}_{3/2}$ .



### 3. Forbidden-Pattern Theory

In chapter 2 we defined the DDH and the STH. The classes of both concatenation hierarchies formalize the famous dot-depth problem in terms of the decidability of their membership problems. The difficulty of finding decision algorithms for the single classes is that one has to find an effective criterion relating the structure of a given DFA with the descriptonal complexity of the accepted language. A certain kind of these criteria are forbidden-pattern characterizations which are results of the following type: “A language  $L$  belongs to the class  $\mathcal{C}$  if and only if the accepting DFA does not have subgraph  $\mathbb{P}$  in its transition graph”. Usually it is easy to test whether a DFA has a certain subgraph  $\mathbb{P}$  which in turn implies the decidability of the class  $\mathcal{C}$ . Moreover, forbidden-pattern characterizations relate the *absence* of a certain subgraph in the DFA with the *existence* of an expression describing the accepted language. So not only is it true that forbidden-pattern characterizations provide decidability, be in fact they even show us the structure (in the DFA) that causes to be a language not in  $\mathcal{C}$ . This antagonism is used in [BKS98] to obtain a gap theorem for leaf-language definable classes between P, NP, and coNP.

So forbidden-pattern characterizations are very powerful characterizations which makes them difficult to find and to prove. Fortunately, such characterizations exist for  $\mathcal{L}_{1/2}$ ,  $\mathcal{L}_1$ ,  $\mathcal{L}_{3/2}$ ,  $\mathcal{B}_{1/2}$  and  $\mathcal{B}_1$  [Sim75, Kna83, Arf91, PW97], and we will prove one for  $\mathcal{B}_{3/2}$ . In this chapter we observe how the patterns that characterize  $\mathcal{L}_{1/2}$  act as building blocks in the patterns characterizing  $\mathcal{L}_{3/2}$ . Surprisingly, we find this observation confirmed, if we compare the pattern for  $\mathcal{B}_{1/2}$  with the characterization of  $\mathcal{B}_{3/2}$  which will be proved in chapter 4. This motivates the introduction of an iteration rule IT on patterns, which continues the observed formation procedure.

In general, starting from an initial class of patterns  $\mathcal{I}$  (fulfilling some reasonable weak assumptions), our iteration rule generates for  $n \geq 0$  classes of patterns  $\mathbb{P}_n^{\mathcal{I}}$  which in turn define language classes  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}})$  if we forbid the patterns  $\mathbb{P}_n^{\mathcal{I}}$  in the transition graphs of DFAs (see Definition 3.5). We prove that  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \cup \text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}}) \cap \text{co}\mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}})$  and, as the main technical result, that  $\text{Pol}(\text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}})) \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}})$  holds (see Theorem 3.16). With the latter we relate in a general way Boolean operations and concatenation to the structural complexity of transition graphs.

We apply our results to particular initial classes of patterns  $\mathcal{B}$  and  $\mathcal{L}$  corresponding to the DDH and the STH, respectively. As a consequence, we obtain strict and decidable hierarchies of classes  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{B}})$  and  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$  which exhaust the class of starfree languages and for which it holds that:

$$\begin{aligned}\mathcal{B}_{n+1/2} &\subseteq \mathcal{FP}(\mathbb{P}_n^{\mathcal{E}}) \\ \mathcal{L}_{n+1/2} &\subseteq \mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})\end{aligned}$$

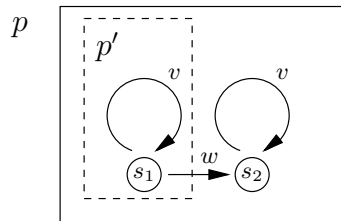
These inclusions imply in particular a lower bound algorithm for the dot-depth of a given language  $L$ . One just has to determine the class  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{E}})$  or  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$  for minimal  $n$  to which  $L$  belongs and it follows that  $L$  has at least dot-depth  $n$ .

It remains to argue that the forbidden-pattern classes are not too large, e.g., if they all equal SF nothing is won. For this end, we provide more structural similarities between the DDH, the STH and the forbidden-pattern classes: All hierarchies show the same inclusion structure (see Figure 3.13) and, interestingly, the typical languages that separate the levels of the DDH and the STH also separate levelwise our forbidden-pattern classes. In particular, it holds that  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$  (just as  $\mathcal{L}_{n+1/2}$ ) does not capture  $\mathcal{B}_{n+1/2}$ .

### 3.1 Looking for an Iteration Rule

Considering the known forbidden-pattern characterizations for  $\mathcal{B}_{1/2}$ ,  $\mathcal{L}_{1/2}$  and  $\mathcal{L}_{3/2}$  from [Arf91, PW97] (cf. Figures 3.1, 3.3 and 3.4), one observes that both states  $s_1$  and  $s_2$  have a loop of the same structure (in case of the pattern for  $\mathcal{L}_{1/2}$  this is an  $\varepsilon$ -loop). We call this the *loop-structure* of the respective pattern.

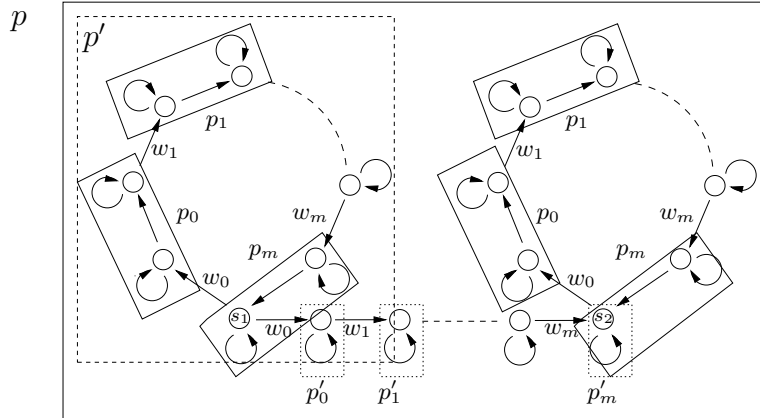
First of all, in Figure 3.1 we recall the class of patterns  $\mathbb{B}_{1/2}$  which characterizes the class of languages  $\mathcal{B}_{1/2}$ . The loop-structure of a pattern  $p \in \mathbb{B}_{1/2}$  is just the  $v$ -loop at  $s_1$  and at  $s_2$ .



**Fig. 3.1.** A pattern  $p \in \mathbb{B}_{1/2}$  with  $v, w \in A^+$  and loop-structure  $p'$ . It is shown in [PW97] that  $\mathbb{B}_{1/2}$  characterizes the class  $\mathcal{B}_{1/2}$ .

In chapter 4 we will prove a forbidden-pattern characterization for  $\mathcal{B}_{3/2}$ . In order to illustrate our iteration rule we anticipate the emerging class of patterns  $\mathbb{B}_{3/2}$  in Figure 3.2. Here the loop-structure  $p'$  of some  $p \in \mathbb{B}_{3/2}$  is more complex: it is the sequence of words  $w_0, w_1, \dots, w_m$  for  $m \geq 0$  such that between each  $w_i, w_{i+1}$  we find some pattern  $p_i$  from  $\mathbb{B}_{1/2}$ . Moreover, we get from  $s_1$  with  $w_0 w_1 \dots w_m$  to  $s_2$  and after each prefix  $w_0 w_1 \dots w_i$  we reach a state with the loop-structure  $p'_i$  (corresponding to the pattern  $p_i$  between  $w_i$  and  $w_{i+1}$ ).

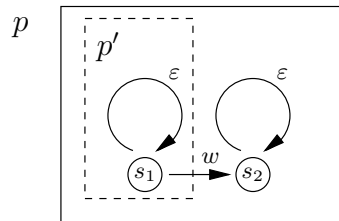
If we generalize the just observed iteration procedure then the next iteration step produces a pattern consisting of two states  $s_1$  and  $s_2$  both having the same loop-structure as follows. There are words  $w_0, w_1, \dots, w_m$  such that between each  $w_i, w_{i+1}$  there is a



**Fig. 3.2.** A pattern  $p \in \mathbb{B}_{3/2}$  with  $w_i \in A^+$ ,  $p_i \in \mathbb{B}_{1/2}$  and loop-structure  $p'$ . In chapter 4 we will prove that  $\mathbb{B}_{3/2}$  characterizes the class  $\mathcal{B}_{3/2}$ .

pattern  $p_i$  now from  $\mathbb{B}_{3/2}$ . Furthermore, going from  $s_1$  to  $s_2$  we should find after every prefix  $w_0w_1 \cdots w_i$  a state with the loop-structure  $p'_i$  of the respective pattern  $p_i \in \mathbb{B}_{3/2}$  that appeared between  $w_i$  and  $w_{i+1}$ . This iteration rule is made precise in the next section.

Surprisingly, we find our iteration rule confirmed if we compare the forbidden-pattern characterizations for  $\mathcal{L}_{1/2}$  and  $\mathcal{L}_{3/2}$ . The class of patterns  $\mathbb{L}_{1/2}$  is given in Figure 3.3. The loop-structure of a pattern  $p \in \mathbb{L}_{1/2}$  is just an  $\varepsilon$ -loop at  $s_1$  and at  $s_2$ .

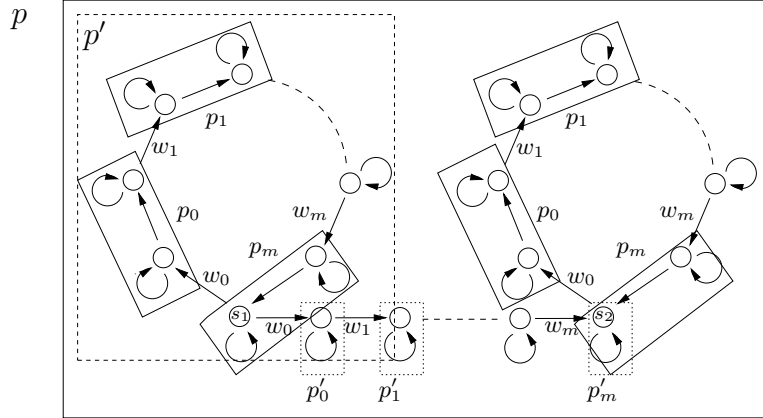


**Fig. 3.3.** A pattern  $p \in \mathbb{L}_{1/2}$  with  $w \in A^*$  and loop-structure  $p'$ . It is shown in [Arf91, PW97] that  $\mathbb{L}_{1/2}$  characterizes the class  $\mathcal{L}_{1/2}$ .

The known forbidden-pattern characterization for  $\mathcal{L}_{3/2}$  [PW97] can be rewritten as the class of patterns  $\mathbb{L}_{3/2}$  which is given in Figure 3.4. In fact, the patterns from  $\mathbb{L}_{3/2}$  looks very similar to those of  $\mathbb{B}_{3/2}$ . The only difference is that for all  $p_i \in \mathbb{L}_{1/2}$  the loop-structure  $p'_i$  is an  $\varepsilon$ -loop.

### 3.2 Forbidden-Pattern Hierarchies

We define the iteration rule IT and provide some useful constructions that let us handle the resulting pattern classes.



**Fig. 3.4.** A pattern  $p \in \mathbb{L}_{3/2}$  with  $w_i \in A^+$ ,  $p_i \in \mathbb{L}_{1/2}$  and loop-structure  $p'$ . Note that the loop-structures of all  $p_i$  are  $\varepsilon$ -loops. From [PW97] it follows that  $\mathbb{L}_{3/2}$  characterizes the class  $\mathcal{L}_{3/2}$ .

### 3.2.1 How to Define Forbidden-Pattern Hierarchies

In order to give an inductive definition for iterated patterns we start with the definition of level 0, i.e., the class of initial patterns. Since this definition should capture the classes of patterns  $\mathbb{B}_{1/2}$  and  $\mathbb{L}_{1/2}$  (cf. Figures 3.1 and 3.3), we define a *class of initial patterns*  $\mathcal{I}$  to be a set of patterns consisting of two states  $s_1$  and  $s_2$  such that (i)  $s_1$  and  $s_2$  have some  $v$ -loop and (ii) some word  $w$  leads from  $s_1$  to  $s_2$ . So in order to describe a single pattern from  $\mathcal{I}$  it suffices to give the pair  $(v, w)$ . Observe that if a DFA has the pattern given in Figure 3.1, then it has also the following “pumped up” patterns:

1. the pattern from Figure 3.1 where  $v, w$  are replaced by  $v^r, wv^r$
2. the pattern from Figure 3.1 where  $v, w$  are replaced by  $v, v$

The second pattern can be found for instance between the states  $s_1, s_1$ . We demand that a class of initial patterns  $\mathcal{I}$  is closed under this kind of pumping transformation. The following definition makes this precise.

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

In Definition 3.5 below we define what it means that a DFA  $\mathcal{M}$  has a pattern from  $\mathbb{P}$ . For this purpose we introduce certain reachability conditions, namely that the loop-structure of some pattern  $p \in \mathbb{P}$  appears at some state and that two states are connected by some pattern  $p \in \mathbb{P}$ . This is consistent with our intuition of finding patterns in transition graphs since the witnessing states are all existentially quantified. As the first step of an inductive definition we consider a class of initial patterns.

**Definition 3.2.** Let  $\mathcal{I}$  be a class of initial patterns. For  $p = (v, w) \in \mathcal{I}$  and given states  $s, s_1, s_2$  of some DFA  $\mathcal{M}$  we say

- $p$  appears at  $s \iff_{\text{def}} s$  has a  $v$ -loop and
- $s_1, s_2$  are connected via  $p$  (in notation  $s_1 \xrightarrow[p]{p} s_2$ )  $\iff_{\text{def}}$   $p$  appears at  $s_1$  and at  $s_2$ , and  $s_1 \xrightarrow{w} s_2$ .

For example we consider the class of patterns  $\mathbb{B}_{1/2}$  which is given in Figure 3.1. Since  $v$  and  $w$  are existentially quantified,  $\mathbb{B}_{1/2}$  can be described in our notations by the class of initial patterns  $\mathcal{B}$  with  $\mathcal{B} = A^+ \times A^+$ . As in Figure 3.1 let  $p = (v, w) \in \mathcal{B}$  for words  $v, w$ . Then  $p$  appears at some state if and only if this state has a  $v$ -loop (which is the loop-structure of pattern  $p$ ). Furthermore, two states  $s_1, s_2$  are connected via  $p$  if and only if both states have a  $v$ -loop and  $w$  leads from  $s_1$  to  $s_2$ . This shows that the notion of connecting two states fits to our intuition of finding a certain structure between these states. Now we formalize the iteration rule.

**Definition 3.3.** For every set  $\mathbb{P}$  we define

$$\text{IT}(\mathbb{P}) =_{\text{def}} \{ (w_0, p_0, \dots, w_m, p_m) \mid m \geq 0, p_i \in \mathbb{P}, w_i \in A^+ \}.$$

If we start with a class of initial patterns  $\mathcal{I}$  and if we apply the iteration rule repeatedly then we obtain a hierarchy of pattern classes. For patterns  $p$  of these classes we have to say what it means that  $p$  connects two states (respectively, appears at a state) of some DFA.

**Definition 3.4.** For a class of initial patterns  $\mathcal{I}$  we set  $\mathbb{P}_0^{\mathcal{I}} =_{\text{def}} \mathcal{I}$  and  $\mathbb{P}_{n+1}^{\mathcal{I}} =_{\text{def}} \text{IT}(\mathbb{P}_n^{\mathcal{I}})$  for  $n \geq 0$ . For  $p = (w_0, p_0, \dots, w_m, p_m) \in \text{IT}(\mathbb{P}_n^{\mathcal{I}})$  and given states  $s, s_1, s_2$  of some DFA  $\mathcal{M}$  we say

- $p$  appears at state  $s \iff_{\text{def}}$  there exist states  $q_0, r_0, \dots, q_m, r_m$  of  $\mathcal{M}$  such that  $s \xrightarrow{w_0} q_0 \xrightarrow[p_0]{p_0} r_0 \xrightarrow{w_1} q_1 \xrightarrow[p_1]{p_1} r_1 \xrightarrow{w_2} \dots \xrightarrow{w_m} q_m \xrightarrow[p_m]{p_m} r_m = s$
- $s_1, s_2$  are connected via  $p$  (in notation  $s_1 \xrightarrow[p]{p} s_2$ )  $\iff_{\text{def}}$   $p$  appears at  $s_1$  and at  $s_2$ , and there exist states  $q_0, \dots, q_m$  of  $\mathcal{M}$  such that  $p_i$  appears at state  $q_i$  for  $0 \leq i \leq m$  and  $s_1 \xrightarrow{w_0} q_0 \xrightarrow{w_1} q_1 \xrightarrow{w_2} \dots \xrightarrow{w_m} q_m = s_2$

Again, let us comment on this definition and see how we can understand it with the known results at hand. Consider the class of initial patterns  $\mathcal{B} = A^+ \times A^+$  and some  $p \in \mathbb{P}_1^{\mathcal{B}}$ . This means that  $p = (w_0, p_0, \dots, w_m, p_m)$  for words  $w_i$  and patterns  $p_i \in \mathbb{P}_0^{\mathcal{B}} = \mathcal{B}$ . The loop-structure described by  $p$  is a loop with factors of words  $w_0, w_1, \dots, w_m$  in this ordering such that between each  $w_i, w_{i+1}$  we find the pattern  $p_i$ . Here we see how elements of  $\mathbb{P}_0^{\mathcal{B}}$  appear as building blocks in the loop-structure of elements of  $\mathbb{P}_1^{\mathcal{B}}$ . The pattern  $p$  connects two states  $s_1, s_2$  if and only if we find the loop-structure of  $p$  at both states and it holds that  $s_1 \xrightarrow{w_1 \dots w_m} s_2$  such that after each prefix  $w_0 \dots w_i$  we reach a state where the loop-structure of  $p_i$  (i.e.,  $p'_i$  in Figure 3.2) appears. Finally, we define in a formal way what it means that some DFA has a pattern from  $\mathbb{P}_n^{\mathcal{I}}$ .

**Definition 3.5.** For a DFA  $\mathcal{M} = (A, S, \delta, s_0, S')$ , a class of initial patterns  $\mathcal{I}$  and  $n \geq 0$  we say that  $\mathcal{M}$  has a pattern from  $\mathbb{P}_n^{\mathcal{I}}$  if and only if there exist states  $s_1, s_2 \in S$ , a word  $z \in A^*$  and a pattern  $p \in \mathbb{P}_n^{\mathcal{I}}$  such that  $s_0 \xrightarrow{+} s_1, s_1 \xrightarrow[p]{p} s_2, s_1 \xrightarrow{z} +$  and  $s_2 \xrightarrow{-} -$ .

### 3.2.2 Transformations of Patterns

To handle patterns  $p \in \mathbb{P}_n^{\mathcal{I}}$  in a better way, we define a word  $\bar{p}^\circ$  obtained from the loop-structure of  $p$  (call this the *loop-word*), and a word  $\bar{p}$  obtained from the subgraph that bridges from  $s_1$  to  $s_2$  (*bridge-word*). With these words we are able to give two useful constructions to obtain from a given pattern  $p \in \mathbb{P}_n^{\mathcal{I}}$  a new pattern from  $\mathbb{P}_n^{\mathcal{I}}$  having certain nice properties.

**Definition 3.6.** *Let  $\mathcal{I}$  be a class of initial patterns. For  $p = (v, w) \in \mathbb{P}_0^{\mathcal{I}}$  we define  $\bar{p} =_{\text{def}} w$  and  $\bar{p}^\circ =_{\text{def}} v$ . For  $n \geq 0$  and  $p = (w_0, p_0, \dots, w_m, p_m) \in \mathbb{P}_{n+1}^{\mathcal{I}}$  we define  $\bar{p} =_{\text{def}} w_0 \cdots w_m$  and  $\bar{p}^\circ =_{\text{def}} w_0 \bar{p}_0 \cdots w_m \bar{p}_m$ .*

In order to establish a relation between the polynomial closure operation and the iteration rule the following two constructions are needed. First, for  $p \in \mathbb{P}_n^{\mathcal{I}}$  some  $\lambda(p) \in \mathbb{P}_n^{\mathcal{I}}$  can be defined such that if  $p$  appears at some state  $s$  then  $s, s$  are connected via  $\lambda(p)$  (see Definition 3.8 and Lemma 3.9). Secondly, in Definition 3.10 we pump up the loop-structure of  $p$  to construct for given  $r \geq 3$  some  $\pi(p, r) \in \mathbb{P}_n^{\mathcal{I}}$  such that:

- if two states are connected via  $p$ , then they are also connected via  $\pi(p, r)$  (see Lemma 3.11)
- in every DFA  $\mathcal{M}$  with  $|\mathcal{M}| \leq r$  the words  $\overline{\pi(p, r)}$  and  $\overline{\pi(p, r)}^\circ$  lead to states where  $\pi(p, r)$  appears (see Lemma 3.12)
- in every DFA  $\mathcal{M}$  with  $|\mathcal{M}| \leq r$  the words  $\overline{\pi(p, r)}^\circ$  and  $\overline{\pi(p, r)}^\circ \overline{\pi(p, r)}$  lead to states that are connected via  $\pi(p, r)$  (see Lemma 3.12)

First of all in the following proposition we state some basic properties of loop-words, bridge-words and patterns  $p \in \mathbb{P}_n^{\mathcal{I}}$ .

**Proposition 3.7.** *Let  $\mathcal{I}$  be a class of initial patterns,  $n \geq 0$ ,  $p \in \mathbb{P}_n^{\mathcal{I}}$  and let  $s, s_1, s_2$  be states of some DFA.*

1. *If  $n \geq 1$  then  $\bar{p}, \bar{p}^\circ \in A^+$ .*
2. *If  $s_1 \xrightarrow{p} s_2$  then  $s_1 \xrightarrow{\bar{p}} s_2$  and  $p$  appears at  $s_1$  and at  $s_2$ .*
3. *If  $p$  appears at state  $s$  then  $s \xrightarrow{\bar{p}^\circ} s$ .*
4. *If  $p$  appears at state  $s$  and if  $p = (w_0, p_0, \dots, w_m, p_m)$  with  $p_i \in \mathbb{P}_{n-1}^{\mathcal{I}}$  for  $n \geq 1$ , then also  $p_m$  appears at state  $s$ .*

*Proof.* All statements are immediate from the definitions. □

We give the construction of  $\lambda(p)$  which connects the states  $s, s$  if  $p$  appears at  $s$ .

**Definition 3.8.** *Let  $\mathcal{I}$  be a class of initial patterns. For  $p = (v, w) \in \mathbb{P}_0^{\mathcal{I}}$  we define  $\lambda(p) =_{\text{def}} (v, v)$ . For  $n \geq 1$  and  $p = (w_0, p_0, \dots, w_m, p_m) \in \mathbb{P}_n^{\mathcal{I}}$  we define  $\lambda(p) =_{\text{def}} (\bar{p}^\circ, \lambda(p_m))$ .*

The following lemma states the announced property of  $\lambda(p)$ .

**Lemma 3.9.** *For every class of initial patterns  $\mathcal{I}$ ,  $n \geq 0$  and  $p \in \mathbb{P}_n^{\mathcal{I}}$  we have  $\lambda(p) \in \mathbb{P}_n^{\mathcal{I}}$ . Moreover, if  $p$  appears at state  $s$  of some DFA then  $s, s$  are connected via  $\lambda(p)$ , i.e.,  $s \xrightarrow{\lambda(p)} s$ .*

*Proof.* We prove the lemma by induction on  $n$ . For  $n = 0$  we have  $p = (v, w) \in \mathbb{P}_0^{\mathcal{I}} = \mathcal{I}$ . By Definition 3.1 it holds that  $\lambda(p) = (v, v) \in \mathcal{I} = \mathbb{P}_0^{\mathcal{I}}$ . If  $p$  appears at some state  $s$  we have  $\delta(s, v) = s$  by definition. Therefore, the states  $s, s$  are connected via  $\lambda(p) = (v, v)$  by Definition 3.2.

Assume the lemma holds for some  $n \geq 0$  and we want to prove it for  $n+1$ . Let  $p \in \mathbb{P}_{n+1}^{\mathcal{I}}$  such that for some  $m \geq 0$ ,  $p_i \in \mathbb{P}_n^{\mathcal{I}}$  and  $w_i \in A^+$  we have  $p = (w_0, p_0, \dots, w_m, p_m)$ . By Proposition 3.7.1 we have  $\bar{p}^\circ \in A^+$  and from induction hypothesis we know that  $\lambda(p_m) \in \mathbb{P}_n^{\mathcal{I}}$ . So with Definition 3.3 we see that  $\lambda(p) = (\bar{p}^\circ, \lambda(p_m)) \in \text{IT}(\mathbb{P}_n^{\mathcal{I}}) = \mathbb{P}_{n+1}^{\mathcal{I}}$ .

It remains to show that the states  $s, s$  are connected via  $\lambda(p) = (\bar{p}^\circ, \lambda(p_m))$  in some DFA if  $p$  appears at state  $s$ . By Proposition 3.7.4 we know that  $p_m$  appears at state  $s$ . So we get from the induction hypothesis and Proposition 3.7.3 that

$$s \xrightarrow{\bar{p}^\circ} s \xrightarrow{\lambda(p_m)} s.$$

It follows that  $\lambda(p)$  appears at state  $s$ . Now let  $s_1 =_{\text{def}} s$ ,  $s_2 =_{\text{def}} s$  and  $q_0 =_{\text{def}} s$ . Then  $q_0 = s_2$  and since  $p$  appears at state  $s$  it follows from Proposition 3.7.3 that  $s_1 \xrightarrow{\bar{p}^\circ} q_0$ . We have already seen that  $s, s$  are connected via  $\lambda(p_m)$ , particularly  $\lambda(p_m)$  appears at state  $s = q_0$  by Proposition 3.7.2. This shows that  $s, s$  are connected via  $\lambda(p)$ .  $\square$

The second construction, i.e., the construction of  $\pi(p, r)$  is more involved.

**Definition 3.10.** Let  $\mathcal{I}$  be a class of initial patterns and  $r \geq 3$ . For  $p = (v, w) \in \mathbb{P}_0^{\mathcal{I}}$  let  $\pi(p, r) =_{\text{def}} (v^{r!}, w \cdot v^{r!})$ . For  $n \geq 1$  and  $p = (w_0, p_0, \dots, w_m, p_m) \in \mathbb{P}_n^{\mathcal{I}}$  we define the following:

$$\begin{aligned} p'_i &=_{\text{def}} \pi(p_i, r) \quad \text{for } 0 \leq i \leq m \\ w &=_{\text{def}} w_0 \cdot \overline{p'_0} \cdot \overline{p'_0} \cdot w_1 \cdot \overline{p'_1} \cdot \overline{p'_1} \cdots w_m \cdot \overline{p'_m} \cdot \overline{p'_m} \\ \pi(p, r) &=_{\text{def}} (w_0 \cdot \overline{p'_0}, p'_0, w_1 \cdot \overline{p'_1}, p'_1, \dots, w_m \cdot \overline{p'_m}, p'_m, \underbrace{w, \lambda(p'_m), \dots, w, \lambda(p'_m)}_{(r-1) \text{ times } "w, \lambda(p'_m)"}) \end{aligned}$$

Next we show that  $\pi(p, r)$  is equivalent to  $p$ , i.e., it appears at a state and connects states in some DFA if  $p$  does.

**Lemma 3.11.** Let  $\mathcal{I}$  be a class of initial patterns,  $r \geq 3$ ,  $n \geq 0$ ,  $p \in \mathbb{P}_n^{\mathcal{I}}$  and let  $s, s_1, s_2$  be states of some DFA.

1. It holds that  $\pi(p, r) \in \mathbb{P}_n^{\mathcal{I}}$ .
2. If  $p$  appears at some state  $s$  then also  $\pi(p, r)$  appears at  $s$ .
3. If  $s_1 \xrightarrow{p} s_2$  then  $s_1 \xrightarrow{\pi(p, r)} s_2$ .

*Proof.* Let  $\mathcal{I}$  be a class of initial patterns and  $r \geq 3$ . We will prove the lemma by induction on  $n$ .

**Induction base:**

For  $n = 0$  and  $p = (v, w) \in \mathbb{P}_0^{\mathcal{I}}$  we have  $\pi(p, r) = (v^{r!}, w \cdot v^{r!})$ . From Definition 3.1 it follows that  $\pi(p, r) \in \mathbb{P}_0^{\mathcal{I}}$ . Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a 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  $\pi(p, r)$  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  $\pi(p, r)$ . 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  $p \in \mathbb{P}_{l+1}^{\mathcal{I}}$ ,  $p' =_{\text{def}} \pi(p, r)$  and choose suitable  $m \geq 0$ ,  $p_i \in \mathbb{P}_l^{\mathcal{I}}$ ,  $w_i \in A^+$  such that  $p = (w_0, p_0, \dots, w_m, p_m)$ . As in Definition 3.10 let  $p'_i =_{\text{def}} \pi(p_i, r)$  and  $w =_{\text{def}} w_0 \cdot \overline{p'_0} \cdot \overline{p'_0} \cdots w_m \cdot \overline{p'_m} \cdot \overline{p'_m}$ . Moreover, let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA and  $s, s_1, s_2 \in S$ . First of all let us show the following claim.

**Claim.** *Let  $s' \in S$  such that  $p$  appears at  $s'$ . Then it holds that  $\delta(s', w) = s'$ .*

If  $p$  appears at  $s'$ , then there exist states  $q_0, r_0, \dots, q_m, r_m$  of  $\mathcal{M}$  such that

$$s' \xrightarrow{w_0} q_0 \xrightarrow{\overline{p_0}} r_0 \xrightarrow{w_1} q_1 \xrightarrow{\overline{p_1}} r_1 \xrightarrow{w_2} \cdots \xrightarrow{w_m} q_m \xrightarrow{\overline{p_m}} r_m = s'$$

By induction hypothesis,  $q_i, r_i$  are also connected via  $p'_i$  for  $0 \leq i \leq m$ . From Proposition 3.7 it follows that  $q_i$  has a  $\overline{p'_i}$ -loop and  $\delta(q_i, \overline{p'_i}) = r_i$  for  $0 \leq i \leq m$ . Hence we have

$$s' \xrightarrow{w_0} q_0 \xrightarrow{\overline{p'_0}} q_0 \xrightarrow{\overline{p'_0}} r_0 \xrightarrow{w_1} q_1 \xrightarrow{\overline{p'_1}} q_1 \xrightarrow{\overline{p'_1}} r_1 \xrightarrow{w_2} \cdots \xrightarrow{w_m} q_m \xrightarrow{\overline{p'_m}} q_m \xrightarrow{\overline{p'_m}} r_m = s'$$

which proves our claim.

By induction hypothesis and Lemma 3.9 we have  $p'_0, \dots, p'_m \in \mathbb{P}_l^{\mathcal{I}}$  and  $\lambda(p'_m) \in \mathbb{P}_l^{\mathcal{I}}$ . Since  $w \in A^+$  and  $w_0 \overline{p'_0}, \dots, w_m \overline{p'_m} \in A^+$ , it follows that  $\pi(p, r) \in \text{IT}(\mathbb{P}_l^{\mathcal{I}}) = \mathbb{P}_{l+1}^{\mathcal{I}}$ . This shows statement 1.

For the proof of statement 2 we assume that  $p$  appears at state  $s$ . Thus there exist states  $q_0, r_0, \dots, q_m, r_m \in S$  such that

$$s \xrightarrow{w_0} q_0 \xrightarrow{\overline{p_0}} r_0 \xrightarrow{w_1} q_1 \xrightarrow{\overline{p_1}} r_1 \xrightarrow{w_2} \cdots \xrightarrow{w_m} q_m \xrightarrow{\overline{p_m}} r_m = s.$$

Using the additional states  $q_j =_{\text{def}} s$  and  $r_j =_{\text{def}} s$  for  $m+1 \leq j \leq m+r!-1$  we will show that also  $p'$  appears at state  $s$ . With  $m' =_{\text{def}} m+r!-1$  we have to show the following.

$$\begin{aligned} s &\xrightarrow{w_0 \overline{p'_0}} q_0 \xrightarrow{\overline{p'_0}} r_0 \xrightarrow{w_1 \overline{p'_1}} q_1 \xrightarrow{\overline{p'_1}} r_1 \xrightarrow{w_2 \overline{p'_2}} \cdots \xrightarrow{w_m \overline{p'_m}} q_m \xrightarrow{\overline{p'_m}} r_m = s \\ s &\xrightarrow{w} q_{m+1} \xrightarrow{\overline{\lambda(p'_m)}} r_{m+1} \xrightarrow{w} \cdots \xrightarrow{w} q_{m'} \xrightarrow{\overline{\lambda(p'_m)}} r_{m'} = s \end{aligned}$$

Since  $q_j \xrightarrow{\overline{p'_j}} r_j$  for  $0 \leq j \leq m$ , we have by induction hypothesis that  $q_j \xrightarrow{\overline{p'_j}} r_j$  for  $0 \leq j \leq m$ . From Proposition 3.7 we obtain that  $p'_j$  appears at  $q_j$  and that  $\delta(q_j, \overline{p'_j}) = q_j$  for  $0 \leq j \leq m$ . This shows

$$s \xrightarrow{w_0 \overline{p'_0}} q_0 \xrightarrow{\overline{p'_0}} r_0 \xrightarrow{w_1 \overline{p'_1}} q_1 \xrightarrow{\overline{p'_1}} r_1 \xrightarrow{w_2 \overline{p'_2}} \cdots \xrightarrow{w_m \overline{p'_m}} q_m \xrightarrow{\overline{p'_m}} r_m = s.$$

Since  $p$  appears at  $s$ , we have by Proposition 3.7.4 that  $p_m$  appears at  $s$ . From Lemma 3.9 we obtain that  $s, s$  are connected via  $\lambda(p_m)$ -loop. Together with our claim from above this shows



$$s \xrightarrow{w} q_{m+1} \xrightarrow{\lambda(p'_m)} r_{m+1} \xrightarrow{w} \cdots \xrightarrow{w} q_{m'} \xrightarrow{\lambda(p'_m)} r_{m'} = s.$$

For statement 3 we assume that 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  $q_0, \dots, q_m \in S$  with

$$s_1 \xrightarrow{w_0} q_0 \xrightarrow{w_1} q_1 \xrightarrow{w_2} \cdots \xrightarrow{w_m} q_m = s_2$$

such that  $p_j$  appears at  $q_j$  for  $0 \leq j \leq m$ . From statement 2 we obtain that also  $p'$  appears at  $s_1$  and at  $s_2$ . So using the additional states  $q_j =_{\text{def}} s_2$  for  $m+1 \leq j \leq m'$  with  $m' =_{\text{def}} m + r! - 1$ , it suffices to show that

$$s_1 \xrightarrow{w_0 \overline{p'_0}} q_0 \xrightarrow{w_1 \overline{p'_1}} q_1 \xrightarrow{w_2 \overline{p'_2}} \cdots \xrightarrow{w_m \overline{p'_m}} q_m \xrightarrow{w} q_{m+1} \cdots \xrightarrow{w} q_{m'} = s_2$$

such that  $p'_j$  appears at  $q_j$  for  $0 \leq j \leq m$ , and  $\lambda(p'_m)$  appears at  $q_j$  for  $m < j \leq m'$ .

Since  $p_j$  appears at  $q_j$  we have by induction hypothesis that also  $p'_j$  appears at  $q_j$  for  $0 \leq j \leq m$ . From Proposition 3.7.3 it follows that  $q_j$  has a  $\overline{p'_j}$ -loop. This shows

$$s_1 \xrightarrow{w_0 \overline{p'_0}} q_0 \xrightarrow{w_1 \overline{p'_1}} q_1 \xrightarrow{w_2 \overline{p'_2}} \cdots \xrightarrow{w_m \overline{p'_m}} q_m = s_2$$

such that  $p'_j$  appears at  $q_j$  for  $0 \leq j \leq m$ . Note that  $s_2$  has a  $w$ -loop which follows from our claim. So it remains to show that  $\lambda(p'_m)$  appears at  $s_2$ . For this we observe that  $p'_m$  appears at  $s_2$  which follows from the induction hypothesis and Proposition 3.7.4. Moreover, from Lemma 3.9 it follows that  $s_2 \xrightarrow{\lambda(p'_m)} s_2$ . Together with Proposition 3.7.2 this yields that  $\lambda(p'_m)$  appears at  $s_2$ .  $\square$

The nice thing about the construction of  $\pi(p, r)$  is that in every DFA of size  $\leq r$  the bridge-word and the loop-word of  $\pi(p, r)$  lead to states where  $\pi(p, r)$  appears. So we have obtained a possibility to find patterns in DFAs for which we only require that their size is  $\leq r$ . Note in particular that we do not require a minimal or permutationfree DFA.

**Lemma 3.12.** *Let  $\mathcal{I}$  be a class of initial patterns,  $r \geq 3$ ,  $n \geq 0$ ,  $p \in \mathbb{P}_n^{\mathcal{I}}$  and let  $\mathcal{M}$  be a DFA with  $|\mathcal{M}| \leq r$ . It holds that*

1.  $\overline{\pi(p, r)}$  leads to states in  $\mathcal{M}$  where  $\pi(p, r)$  appears,
2.  $\overline{\pi(p, r)}$  leads to states in  $\mathcal{M}$  where  $\pi(p, r)$  appears,
3.  $\overline{\pi(p, r)}$  and  $\overline{\pi(p, r)} \overline{\pi(p, r)}$  lead to states in  $\mathcal{M}$  that are connected via  $\pi(p, r)$  and
4.  $\overline{\pi(p, r)}$  and  $\overline{\pi(p, r)} \overline{\pi(p, r)}$  lead to states in  $\mathcal{M}$  that are connected via  $\pi(p, r)$ .

*Proof.* We prove the lemma by induction on  $n$ .

**Induction base.** Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be some DFA with  $|\mathcal{M}| \leq r$ . If  $n = 0$  then we have  $p = (v, w) \in \mathbb{P}_0^{\mathcal{I}}$  and  $\pi(p, r) = (v^r!, w \cdot v^r!)$ . Since  $v^r$  leads to  $v^r!$ -loops in  $\mathcal{M}$  (Proposition 1.1) we obtain that  $\overline{\pi(p, r)} = v^{r!-r} \cdot v^r$  and  $\overline{\pi(p, r)} = w \cdot v^{r!-r} \cdot v^r$  lead to states where  $\pi(p, r)$  appears. Hence  $\overline{\pi(p, r)}$  and  $\overline{\pi(p, r)} \overline{\pi(p, r)}$  lead to states which are connected via  $\pi(p, r)$ , and the same holds also for  $\overline{\pi(p, r)}$  and  $\overline{\pi(p, r)} \overline{\pi(p, r)}$ .

**Induction step.** Suppose we have shown the lemma for some  $n \geq 0$  and we want to show it for  $n+1$ . Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be some DFA with  $|\mathcal{M}| \leq r$ . Furthermore, let  $p = (w_0, p_0, \dots, w_m, p_m) \in \mathbb{P}_{n+1}^{\mathcal{I}}$  and assume that  $w, p'_i$  are as in Definition 3.10. First we show the following claim.

**Claim.** *It holds that  $w^{r!-1}$  leads to states in  $\mathcal{M}$  where  $\pi(p, r)$  appears.*

Observe that  $w^{r!-1} = w^{r!-1-r}w^r$  leads to a  $w^{r!}$ -loop in  $\mathcal{M}$  by Proposition 1.1 since  $r! \geq r+1$  for  $r \geq 3$ . So let  $s$  be a state of  $\mathcal{M}$  that has a  $w^{r!}$ -loop, we will show that  $\pi(p, r)$  appears at  $s$ . Let  $m' =_{\text{def}} m + r! - 1$  and define the witnessing states  $q_0, r_0, \dots, q_{m'}, r_{m'}$  as follows.

$$\begin{aligned} s &\xrightarrow{w_0 \cdot \overline{p'_0}} q_0 \xrightarrow{\overline{p'_0}} r_0 \xrightarrow{w_1 \cdot \overline{p'_1}} q_1 \xrightarrow{\overline{p'_1}} r_1 \xrightarrow{w_2 \cdot \overline{p'_2}} \dots \xrightarrow{w_m \cdot \overline{p'_m}} q_m \xrightarrow{\overline{p'_m}} r_m \\ r_m &\xrightarrow{w} q_{m+1} = r_{m+1} \xrightarrow{w} q_{m+2} = r_{m+2} \xrightarrow{w} \dots \xrightarrow{w} q_{m'} = r_{m'} \end{aligned}$$

It follows from the induction hypothesis that  $q_i, r_i$  are connected via  $p'_i$  for  $0 \leq i \leq m$ . Moreover, the hypothesis also shows that  $p'_m$  appears at  $q_j$  for  $m+1 \leq j \leq m'$  since  $\overline{p'_m}$  is a suffix of  $w$ . From Lemma 3.9 we get that  $q_j, r_j$  are connected via  $\lambda(p'_m)$ . Finally, by the definition of  $w$  we have  $r_m = \delta(s, w)$  and  $r_{m'} = \delta(r_m, w^{r!-1}) = \delta(s, w^{r!}) = s$ . Hence we have shown that

$$\begin{aligned} s &\xrightarrow{w_0 \cdot \overline{p'_0}} q_0 \xrightarrow{\overline{p'_0}} r_0 \xrightarrow{w_1 \cdot \overline{p'_1}} q_1 \xrightarrow{\overline{p'_1}} r_1 \xrightarrow{w_2 \cdot \overline{p'_2}} \dots \xrightarrow{w_m \cdot \overline{p'_m}} q_m \xrightarrow{\overline{p'_m}} r_m \quad \text{and} \\ r_m &\xrightarrow{w} q_{m+1} \xrightarrow{\lambda(p'_m)} r_{m+1} \xrightarrow{w} q_{m+2} \xrightarrow{\lambda(p'_m)} r_{m+2} \xrightarrow{w} \dots \xrightarrow{w} q_{m'} \xrightarrow{\lambda(p'_m)} r_{m'} = s. \end{aligned}$$

So  $\pi(p, r)$  appears at  $s$  which completes the proof of our claim.

We come to the proof of the statements 1 and 2. Since  $\overline{p'_m}$  is a suffix of  $w$  it follows from the induction hypothesis that  $w$  leads to states where  $p'_m$  appears. From Lemma 3.9 and Proposition 3.7.2 we obtain that  $w$  leads to a  $\overline{\lambda(p'_m)}$ -loop in  $\mathcal{M}$ . Hence our claim holds also for  $(w \cdot \overline{\lambda(p'_m)})^{r!-1}$  instead of  $w$ . By definition we have

$$\begin{aligned} \overline{\pi(p, r)} &= w_0 \cdot \overline{p'_0} \dots w_m \cdot \overline{p'_m} \cdot w^{r!-1} \quad \text{and} \\ \overline{\pi(p, r)}^\circ &= w_0 \cdot \overline{p'_0} \cdot \overline{p'_0} \dots w_m \cdot \overline{p'_m} \cdot \overline{p'_m} \cdot (w \cdot \overline{\lambda(p'_m)})^{r!-1}. \end{aligned}$$

So the claim says that  $\overline{\pi(p, r)}$  and  $\overline{\pi(p, r)}^\circ$  lead to states in  $\mathcal{M}$  where  $\pi(p, r)$  appears. This shows the statements 1 and 2 of the lemma.

We turn to statement 3 and choose an arbitrary state  $s$  of  $\mathcal{M}$ . For  $s_1 =_{\text{def}} \delta(s, \overline{\pi(p, r)}^\circ)$  and  $s_2 =_{\text{def}} \delta(s, \overline{\pi(p, r)} \cdot \overline{\pi(p, r)})$  we show that  $s_1, s_2$  are connected via  $\pi(p, r)$ . Again let  $m' =_{\text{def}} m + r! - 1$  and define the witnessing states  $q_0, \dots, q_{m'}$  as follows.

$$s_1 \xrightarrow{w_0 \cdot \overline{p'_0}} q_0 \xrightarrow{w_1 \cdot \overline{p'_1}} q_1 \xrightarrow{w_2 \cdot \overline{p'_2}} \dots \xrightarrow{w_m \cdot \overline{p'_m}} q_m \xrightarrow{w} q_{m+1} \xrightarrow{w} \dots \xrightarrow{w} q_{m'}$$

We have already seen in the proof of the statements 1 and 2 that  $\pi(p, r)$  appears at  $s_1$  and at  $s_2$ . Observe that  $q_{m'} = \delta(s_1, \overline{\pi(p, r)}) = s_2$ . So it remains to show that  $p'_i$  appears at  $q_i$  for  $0 \leq i \leq m$  and that  $\lambda(p'_m)$  appears at  $q_j$  for  $m+1 \leq j \leq m'$ .

By induction hypothesis we have that  $\overline{p'_i}$  leads to states in  $\mathcal{M}$  where  $p'_i$  appears. Hence  $p'_i$  appears at state  $q_i$  for  $0 \leq i \leq m$ . Since  $\overline{p'_m}$  is a suffix of  $w$  the induction hypothesis shows that  $p'_m$  appears at  $q_j$  for all  $j$  with  $m+1 \leq j \leq m'$ . With Lemma 3.9 we see that  $q_j, q_j$  are connected via  $\lambda(p'_m)$ , so in particular  $\lambda(p'_m)$  appears at state  $q_j$ . This shows statement 3. Analogously we prove statement 4, for this we only have to replace the terms  $\overline{\pi(p, r)}^\circ$  by  $\overline{\pi(p, r)}$ .  $\square$

### 3.3 Pattern Iterator versus Polynomial Closure

We relate in this section in a general way Boolean operations and concatenation to the structural complexity of transition graphs. More precisely in Theorem 3.16 we show that a complementation followed by a polynomial closure operation on the language side is captured by our iteration rule on the pattern side.

With Lemma 3.15 we isolate the main argument of the proof of Theorem 3.16. It says that under certain assumptions we can replace bridge-words by their respective loop-words without leaving the language of some DFA.

First of all we use our pattern classes in a forbidden-pattern sense to define language classes which we call *forbidden-pattern classes*. In particular, this implies that each class of initial patterns induces a hierarchy of forbidden-pattern classes. In Proposition 3.14 we show that these classes are well-defined.

**Definition 3.13.** *Let  $\mathcal{I}$  be a class of initial patterns and  $n \geq 0$ .*

$$\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) =_{\text{def}} \{L \subseteq A^+ \mid L = L(\mathcal{M}) \text{ for a DFA } \mathcal{M} \text{ that does not have a pattern from } \mathbb{P}_n^{\mathcal{I}}\}$$

**Proposition 3.14.** *Let  $\mathcal{I}$  be a class of initial patterns,  $n \geq 0$  and let  $\mathcal{M}_1$  and  $\mathcal{M}_2$  be two DFAs such that  $L(\mathcal{M}_1) = L(\mathcal{M}_2)$ . Then it holds that  $\mathcal{M}_1$  has a pattern from  $\mathbb{P}_n^{\mathcal{I}}$  if and only if  $\mathcal{M}_2$  has a pattern from  $\mathbb{P}_n^{\mathcal{I}}$ .*

*Proof.* It suffices to show one implication. So suppose  $\mathcal{M}_1$  has a pattern from  $\mathbb{P}_n^{\mathcal{I}}$ , and denote by  $s_0$  the starting state of  $\mathcal{M}_1$ . Then there are states  $s_1, s_2$  in  $\mathcal{M}_1$ , a word  $z \in A^*$  and some  $p \in \mathbb{P}_n^{\mathcal{I}}$  such that the following holds in  $\mathcal{M}_1$ .

$$s_0 \longrightarrow s_1, \quad s_1 \xrightarrow[p]{\circ} s_2, \quad s_1 \xrightarrow{z} +, \quad s_2 \xrightarrow{z} -$$

Let  $r =_{\text{def}} |\mathcal{M}_2|$  and  $p' =_{\text{def}} \pi(p, r)$ . We obtain from Lemma 3.11 that  $s_1, s_2$  in  $\mathcal{M}_1$  are also connected via  $p' \in \mathbb{P}_n^{\mathcal{I}}$ . So by Proposition 3.7 we have  $x\overline{p'}z \in L(\mathcal{M}_1) = L(\mathcal{M}_2)$  and  $x\overline{p'}\overline{p'}z \notin L(\mathcal{M}_1) = L(\mathcal{M}_2)$ . Denote by  $s'_0$  the initial state of  $\mathcal{M}_2$  and define states  $s'_1$  and  $s'_2$  in  $\mathcal{M}_2$  as follows.

$$s'_0 \xrightarrow{x\overline{p'}} s'_1, \quad s'_0 \xrightarrow{x\overline{p'}\overline{p'}} s'_2$$

By Lemma 3.12.3 we obtain that  $s'_1$  and  $s'_2$  are connected via  $p'$  in  $\mathcal{M}_2$ . Since  $s'_1 \xrightarrow{z} +$  and  $s'_2 \xrightarrow{z} -$  this shows that  $\mathcal{M}_2$  has a pattern from  $\mathbb{P}_n^{\mathcal{I}}$ .  $\square$

**Lemma 3.15.** *Let  $\mathcal{I}$  be a class of initial patterns,  $n \geq 0$ ,  $r \geq 3$ ,  $p \in \mathbb{P}_{n+1}^{\mathcal{I}}$  and  $p' =_{\text{def}} \pi(p, r)$ . Moreover, let  $\mathcal{M} = (A, S, \delta, s_0, S')$  with  $|\mathcal{M}| \leq r$  be a DFA which does not have a pattern from  $\mathbb{P}_n^{\mathcal{I}}$ . Then for all  $x, z \in A^*$  we have*

$$\overline{x\overline{p'}}z \in L(\mathcal{M}) \implies x\overline{p'}z \in L(\mathcal{M}).$$

*Proof.* We choose suitable  $m \geq 0$ ,  $w_0, \dots, w_m \in A^+$  and  $p_0, \dots, p_m \in \mathbb{P}_n^{\mathcal{I}}$  such that  $p = (w_0, p_0, \dots, w_m, p_m)$ . Let  $p'_i$  and  $w$  be as in Definition 3.10, i.e., for  $0 \leq i \leq m$  let  $p'_i =_{\text{def}} \pi(p_i, r)$  and  $w =_{\text{def}} w_0 \cdot \overline{p'_0} \cdot \overline{p'_0} \cdots w_m \cdot \overline{p'_m} \cdot \overline{p'_m}$ . From Definition 3.10 it follows that

$p' = (w_0 \cdot \overline{p'_0}, p'_0, \dots, w_m \cdot \overline{p'_m}, p'_m, w, \lambda(p'_m), \dots, w, \lambda(p'_m))$  where the term “ $w, \lambda(p'_m)$ ” appears  $(r! - 1)$  times. Let  $x, z \in A^*$  such that  $x\overline{p'_m}z \in L(\mathcal{M})$ . Thus we have

$$\underbrace{xw_0 \overline{p'_0}}_{x'=\text{def}} \underbrace{w_1 \overline{p'_1} w_2 \overline{p'_2} \cdots w_m \overline{p'_m} w^{r!-1} z}_{z'=\text{def}} \in L(\mathcal{M}).$$

We want to show that  $x'\overline{p'_0}z' \in L(\mathcal{M})$ . From Lemma 3.12.3 it follows that the states  $s_1 =_{\text{def}} \delta(s_0, x'\overline{p'_0})$  and  $s_2 =_{\text{def}} \delta(s_0, x'\overline{p'_0}z')$  are connected via  $p'_0$ . Note that  $p'_0 \in \mathbb{P}_n^{\mathcal{I}}$  by Lemma 3.11.1. If  $x'\overline{p'_0}z' \notin L(\mathcal{M})$  then we have the following in  $\mathcal{M}$ .

$$s_0 \xrightarrow{x'\overline{p'_0}} s_1, \quad s_1 \xrightarrow{p'_0} s_2, \quad s_1 \xrightarrow{z'} +, \quad s_2 \xrightarrow{z'} -$$

Hence  $\mathcal{M}$  has a pattern from  $\mathbb{P}_n^{\mathcal{I}}$  which is a contradiction to the assumption. Thus starting from

$$xw_0 \overline{p'_0} w_1 \overline{p'_1} w_2 \overline{p'_2} \cdots w_m \overline{p'_m} w^{r!-1} z \in L(\mathcal{M})$$

we have shown

$$xw_0 \overline{p'_0} p'_0 w_1 \overline{p'_1} w_2 \overline{p'_2} \cdots w_m \overline{p'_m} w^{r!-1} z \in L(\mathcal{M}).$$

Analogously we obtain:

$$\begin{aligned} & xw_0 \overline{p'_0} p'_0 w_1 \overline{p'_1} w_2 \overline{p'_2} w_3 \overline{p'_3} w_4 \overline{p'_4} \cdots w_m \overline{p'_m} w^{r!-1} z \in L(\mathcal{M}) \\ & xw_0 \overline{p'_0} p'_0 w_1 \overline{p'_1} p'_1 w_2 \overline{p'_2} w_3 \overline{p'_3} w_4 \overline{p'_4} \cdots w_m \overline{p'_m} w^{r!-1} z \in L(\mathcal{M}) \\ & xw_0 \overline{p'_0} p'_0 w_1 \overline{p'_1} p'_1 w_2 \overline{p'_2} p'_2 w_3 \overline{p'_3} w_4 \overline{p'_4} \cdots w_m \overline{p'_m} w^{r!-1} z \in L(\mathcal{M}) \\ & xw_0 \overline{p'_0} p'_0 w_1 \overline{p'_1} p'_1 w_2 \overline{p'_2} p'_2 w_3 \overline{p'_3} p'_3 w_4 \overline{p'_4} \cdots w_m \overline{p'_m} w^{r!-1} z \in L(\mathcal{M}) \\ & \vdots \\ & xw_0 \overline{p'_0} p'_0 w_1 \overline{p'_1} p'_1 w_2 \overline{p'_2} p'_2 w_3 \overline{p'_3} p'_3 w_4 \overline{p'_4} \cdots w_m \overline{p'_m} p'_m w^{r!-1} z \in L(\mathcal{M}) \end{aligned} \quad (3.1)$$

By definition,  $\overline{p'_m}$  is a suffix of  $w$ . From Lemma 3.12.2 it follows that  $w$  leads to states in  $\mathcal{M}$  where  $p'_m$  appears. Together with Lemma 3.9 we obtain that for all  $s' \in S$  with  $s =_{\text{def}} \delta(s', w)$  it holds that  $s, s'$  are connected via  $\lambda(p'_m)$ . Now from Proposition 3.7.2 it follows that  $w$  leads to states in  $\mathcal{M}$  that have a  $\lambda(p'_m)$ -loop. Thus from equation (3.1) we obtain

$$\underbrace{xw_0 \overline{p'_0} p'_0 w_1 \overline{p'_1} p'_1 w_2 \overline{p'_2} p'_2 w_3 \overline{p'_3} p'_3 w_4 \overline{p'_4} p'_4 \cdots w_m \overline{p'_m} p'_m}_{=\overline{p'^r}} \left( w \lambda(p'_m) \right)^{r!-1} z \in L(\mathcal{M}).$$

This proves the lemma.  $\square$

We come to the main result of this chapter, i.e., the theorem connecting the polynomial closure operation and our iteration rule.

**Theorem 3.16.** *Let  $\mathcal{I}$  be a class of initial patterns and let  $n \geq 0$ .*

$$\text{Pol}(\text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}})) \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}})$$

*Proof.* We assume that there is an  $L \in \text{Pol}(\text{co}\mathcal{FP}(\mathbb{P}_n^x))$  which is not in  $\mathcal{FP}(\mathbb{P}_{n+1}^x)$  and show that this leads to a contradiction. From  $L \in \text{Pol}(\text{co}\mathcal{FP}(\mathbb{P}_n^x))$  it follows that

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

for  $k, k_i \geq 0$  and  $L_{i,j} \in \text{co}\mathcal{FP}(\mathbb{P}_n^x)$ . Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA with  $L(\mathcal{M}) = L$ . For  $1 \leq i \leq k$  and  $0 \leq j \leq k_i$  let  $\mathcal{M}_{i,j}$  be a DFA with  $L(\mathcal{M}_{i,j}) = L_{i,j}$  and let  $\mathcal{M}'_{i,j}$  be a DFA with  $L(\mathcal{M}'_{i,j}) = A^+ \setminus L_{i,j}$ . Furthermore, in order to choose  $r$  sufficiently large we define

$$r =_{\text{def}} \max(\{|\mathcal{M}_{i,j}|, |\mathcal{M}'_{i,j}| \mid 1 \leq i \leq k, 0 \leq j \leq k_i\} \cup \{|\mathcal{M}|, 3\} \cup \{k_i + 1 \mid 1 \leq i \leq k\}).$$

The DFA  $\mathcal{M}$  has a pattern from  $\mathbb{P}_{n+1}^x$  since  $L \notin \mathcal{FP}(\mathbb{P}_{n+1}^x)$  by assumption. So there exist states  $s_1, s_2 \in S$ , words  $x, z \in A^*$  and some  $p \in \mathbb{P}_{n+1}^x$  such that

$$s_0 \xrightarrow{x} s_1, \quad s_1 \xrightarrow{p} s_2, \quad s_1 \xrightarrow{z} + \quad \text{and} \quad s_2 \xrightarrow{z} -.$$

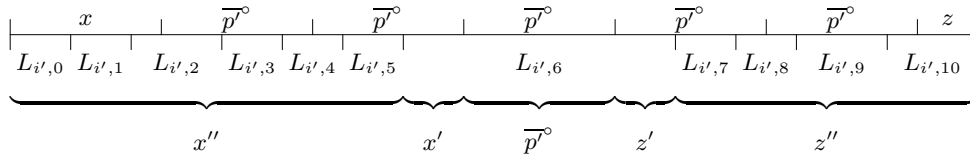
It follows that  $L \neq \emptyset$  and  $k \geq 1$ . By Lemma 3.11 the states  $s_1, s_2$  are also connected via pattern  $p' =_{\text{def}} \pi(p, r)$ . From Proposition 3.7 it follows that  $x(\overline{p'})^i z \in L$  for all  $i \geq 0$ . Thus there exists an  $i'$  with  $1 \leq i' \leq k$  such that

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

Since  $r \geq k_{i'} + 1$ , it follows from a pigeon hole argument that if we decompose the word  $x(\overline{p'})^r z$  with respect to  $L_{i',0}L_{i',1} \cdots L_{i',k_{i'}}$  then there is at least one language  $L_{i',j'}$  whose corresponding factor is of the form  $x'\overline{p'}z'$ . In other words, there exist some  $j'$  with  $0 \leq j' \leq k_{i'}$  and words  $x', x'', z', z'' \in A^*$  such that the following holds.

1. The word  $x(\overline{p'})^r z$  can be decomposed as  $x(\overline{p'})^r z = x''x'\overline{p'}z'z''$ .
2.  $x''x' = x(\overline{p'})^i$  and  $z'z'' = (\overline{p'})^j z$  for some  $i, j \geq 0$ .
3.  $x'' \in L_{i',0}L_{i',1} \cdots L_{i',j'-1}$ ,  $x'\overline{p'}z' \in L_{i',j'}$  and  $z'' \in L_{i',j'+1}L_{i',j'+2} \cdots L_{i',k_{i'}}$ .

An example for this decomposition is shown in Figure 3.5.



**Fig. 3.5.** Decomposition of the word  $x(\overline{p'})^5 z \in L_{i',0}L_{i',1} \cdots L_{i',10}$ .

Since  $|\mathcal{M}'_{i',j'}| \leq r$  the word  $\overline{p'}$  leads to states in  $\mathcal{M}'_{i',j'}$  where  $p'$  appears by Lemma 3.12. In particular, such a state has a  $\overline{p'}$ -loop by Proposition 3.7.3. From  $x'\overline{p'}z' \in L_{i',j'}$  it follows that for all  $i \geq 1$  we have

$$x'(\overline{p'})^i z' \in L_{i',j'}. \quad (3.2)$$

Because  $s_1, s_2$  are connected via  $p'$  in  $\mathcal{M}$  we have by Proposition 3.7 that  $\delta(s_0, x''x') = s_1$ ,  $\delta(s_2, z'z'') = \delta(s_2, z)$  and  $\delta(s_1, \overline{p'}) = s_2$ . Assume that  $x'\overline{p'}\overline{p'}z' \in L_{i',j'}$ . Then we obtain  $x''x'\overline{p'}\overline{p'}z'z'' \in L$ . It follows that

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

This is a contradiction since  $\delta(s_2, z) \notin S'$ . So we have seen that

$$x'\overline{p'}\overline{p'}z' \notin L_{i',j'}.$$

In other terms, it holds that

$$x'\overline{p'}\overline{p'}z' \in A^+ \setminus L_{i',j'} = L(\mathcal{M}'_{i',j'}) \quad (3.3)$$

because  $|x'\overline{p'}\overline{p'}z'| \geq |x'\overline{p'}z'| \geq 1$  (Proposition 3.7.1). Recall that  $L(\mathcal{M}'_{i',j'}) \in \mathcal{FP}(\mathbb{P}_n^{\mathcal{I}})$  and hence the DFA  $\mathcal{M}'_{i',j'}$  does not have a pattern from  $\mathbb{P}_n^{\mathcal{I}}$ . Since  $|\mathcal{M}'_{i',j'}| \leq r$  we can apply Lemma 3.15, and together with (3.3) we obtain  $x'\overline{p'}\overline{p'}z' \in L(\mathcal{M}'_{i',j'})$ . It follows that  $x'\overline{p'}\overline{p'}z' \notin A^+ \setminus L(\mathcal{M}'_{i',j'}) = L_{i',j'}$ . This is a contradiction to (3.2).  $\square$

### 3.4 Inclusion Structure of the Forbidden-Pattern Classes $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}})$

In this section we define a relation  $\triangleleft$  between pattern classes. If two pattern classes are in this relation then this means that every pattern from the second class can be interpreted as a pattern from the first class. This is made precise in Definition 3.17. The main result of this section is stated in Theorem 3.21. It says that for any class of initial patterns  $\mathcal{I}$  satisfying the weak assumption  $\mathbb{P}_0^{\mathcal{I}} \triangleleft \mathbb{P}_1^{\mathcal{I}}$  the corresponding language classes form a hierarchy which shows the same inclusion structure as the DDH and the STH.

**Definition 3.17.** *For classes of initial patterns  $\mathcal{I}_1, \mathcal{I}_2$  and  $n_1, n_2 \geq 0$  we define  $\mathbb{P}_{n_1}^{\mathcal{I}_1} \triangleleft \mathbb{P}_{n_2}^{\mathcal{I}_2}$  if and only if for every  $p_2 \in \mathbb{P}_{n_2}^{\mathcal{I}_2}$  there exists a  $p_1 \in \mathbb{P}_{n_1}^{\mathcal{I}_1}$  such that for every DFA  $\mathcal{M}$  and all states  $s, s_1, s_2$  of  $\mathcal{M}$  the following holds.*

1. If  $p_2$  appears at  $s$ , then  $p_1$  appears at  $s$ .
2. If  $s_1 \xrightarrow{\overline{p_2}} s_2$  then  $s_1 \xrightarrow{\overline{p_1}} s_2$ .

First of all let us prove that our iterator IT respects the relation  $\triangleleft$ .

**Proposition 3.18.** *For classes of initial patterns  $\mathcal{I}_1, \mathcal{I}_2$  and  $n_1, n_2 \geq 0$  the following holds.*

$$\mathbb{P}_{n_1}^{\mathcal{I}_1} \triangleleft \mathbb{P}_{n_2}^{\mathcal{I}_2} \implies \mathbb{P}_{n_1+1}^{\mathcal{I}_1} \triangleleft \mathbb{P}_{n_2+1}^{\mathcal{I}_2}$$

*Proof.* Let  $\mathcal{I}_1, \mathcal{I}_2$  be classes of initial patterns and  $n_1, n_2 \geq 0$  such that  $\mathbb{P}_{n_1}^{\mathcal{I}_1} \triangleleft \mathbb{P}_{n_2}^{\mathcal{I}_2}$ . Hence for a given  $p_2 = (w_{2,0}, p_{2,0}, \dots, w_{2,m}, p_{2,m}) \in \mathbb{P}_{n_2+1}^{\mathcal{I}_2}$  there exist  $p_{1,0}, \dots, p_{1,m} \in \mathbb{P}_{n_1}^{\mathcal{I}_1}$  such that for every DFA  $\mathcal{M}$  and all states  $s, s_1, s_2$  of  $\mathcal{M}$  the following holds.

- (a) If  $p_{2,i}$  appears at  $s$ , then  $p_{1,i}$  appears at  $s$  for  $0 \leq i \leq m$ .  
 (b) If  $s_1 \xrightarrow{\overset{p_{2,i}}{\sigma\sigma\sigma}} s_2$  then  $s_1 \xrightarrow{\overset{p_{1,i}}{\sigma\sigma\sigma}} s_2$  for  $0 \leq i \leq m$ .

We define  $p_1 =_{\text{def}} (w_{2,0}, p_{1,0}, \dots, w_{2,m}, p_{1,m})$  and we observe that  $p_1 \in \mathbb{P}_{n_1+1}^{\mathcal{I}_1}$ . Now let  $\mathcal{M}$  be a DFA and let  $s, s_1, s_2$  be states of  $\mathcal{M}$ . We want to show the following.

- (i) If  $p_2$  appears at  $s$ , then  $p_1$  appears at  $s$ .  
 (ii) If  $s_1 \xrightarrow{\overset{p_2}{\sigma\sigma\sigma}} s_2$  then  $s_1 \xrightarrow{\overset{p_1}{\sigma\sigma\sigma}} s_2$ .

Suppose that  $p_2$  appears at  $s$ , then there are states  $q_0, r_0, \dots, q_m, r_m \in S$  such that

$$s \xrightarrow{w_{2,0}} q_0 \xrightarrow{\overset{p_{2,0}}{\sigma\sigma\sigma}} r_0 \xrightarrow{w_{2,1}} q_1 \xrightarrow{\overset{p_{2,1}}{\sigma\sigma\sigma}} r_1 \xrightarrow{w_{2,2}} \dots \xrightarrow{w_{2,m}} q_m \xrightarrow{\overset{p_{2,m}}{\sigma\sigma\sigma}} r_m = s.$$

From (b) it follows that  $q_i \xrightarrow{\overset{p_{1,i}}{\sigma\sigma\sigma}} r_i$  for  $0 \leq i \leq m$ . Therefore,  $p_1$  appears at  $s$ , and we have shown statement (i).

Suppose now that  $s_1 \xrightarrow{\overset{p_2}{\sigma\sigma\sigma}} s_2$ . By definition,  $p_2$  appears at the states  $s_1$  and  $s_2$ , and there exist states  $q_0, \dots, q_m \in S$  such that  $p_{2,i}$  appears at state  $q_i$  for  $0 \leq i \leq m$  and

$$s_1 \xrightarrow{w_{2,0}} q_0 \xrightarrow{w_{2,1}} q_1 \xrightarrow{w_{2,2}} \dots \xrightarrow{w_{2,m}} q_m = s_2.$$

From statement (i) we obtain that  $p_1$  appears at state  $s_1$  and at state  $s_2$ . Furthermore, from (a) it follows that also  $p_{1,i}$  appears at state  $r_i$  for  $0 \leq i \leq m$ . Hence  $s_1 \xrightarrow{\overset{p_1}{\sigma\sigma\sigma}} s_2$  and we have shown statement (ii). It follows that  $\mathbb{P}_{n_1+1}^{\mathcal{I}_1} \trianglelefteq \mathbb{P}_{n_2+1}^{\mathcal{I}_2}$ .  $\square$

Now it is easy to see that the relation  $\trianglelefteq$  on the pattern side implies inclusion on the language side.

**Proposition 3.19.** *For classes of initial patterns  $\mathcal{I}_1, \mathcal{I}_2$  and  $n_1, n_2 \geq 0$  the following holds.*

$$\mathbb{P}_{n_1}^{\mathcal{I}_1} \trianglelefteq \mathbb{P}_{n_2}^{\mathcal{I}_2} \implies \mathcal{FP}(\mathbb{P}_{n_1}^{\mathcal{I}_1}) \subseteq \mathcal{FP}(\mathbb{P}_{n_2}^{\mathcal{I}_2})$$

*Proof.* Let  $\mathcal{I}_1, \mathcal{I}_2$  be classes of initial patterns and  $n_1, n_2 \geq 0$  such that  $\mathbb{P}_{n_1}^{\mathcal{I}_1} \trianglelefteq \mathbb{P}_{n_2}^{\mathcal{I}_2}$ . For an arbitrary language  $L \subseteq A^+$  with  $L \notin \mathcal{FP}(\mathbb{P}_{n_2}^{\mathcal{I}_2})$  we want to show that  $L \notin \mathcal{FP}(\mathbb{P}_{n_1}^{\mathcal{I}_1})$ . Let  $\mathcal{M}$  be a DFA with  $L(\mathcal{M}) = L \notin \mathcal{FP}(\mathbb{P}_{n_2}^{\mathcal{I}_2})$ . Hence  $\mathcal{M}$  has a pattern from  $\mathbb{P}_{n_2}^{\mathcal{I}_2}$ . This means that there exist states  $s_1, s_2 \in S$ , a word  $z \in A^*$  and some  $p_2 \in \mathbb{P}_{n_2}^{\mathcal{I}_2}$  such that

$$s_0 \longrightarrow s_1, \quad s_1 \xrightarrow{\overset{p_2}{\sigma\sigma\sigma}} s_2, \quad s_1 \xrightarrow{z} +, \quad s_2 \xrightarrow{z} -.$$

Since  $\mathbb{P}_{n_1}^{\mathcal{I}_1} \trianglelefteq \mathbb{P}_{n_2}^{\mathcal{I}_2}$ , there exists a  $p_1 \in \mathbb{P}_{n_1}^{\mathcal{I}_1}$  such that  $s_1 \xrightarrow{\overset{p_1}{\sigma\sigma\sigma}} s_2$ . It follows that  $\mathcal{M}$  has also a pattern from  $\mathbb{P}_{n_1}^{\mathcal{I}_1}$ . This shows  $L \notin \mathcal{FP}(\mathbb{P}_{n_1}^{\mathcal{I}_1})$ .  $\square$

**Proposition 3.20.** *For a class of initial patterns  $\mathcal{I}$  and  $n \geq 0$  the following holds.*

1.  $\text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}})$
2.  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{co}\mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}})$

*Proof.* From Theorem 3.16 it follows that  $\text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{Pol}(\text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}})) \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}})$ . This also implies  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) = \text{co}(\text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}})) \subseteq \text{co}\mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}})$ .  $\square$

The following theorem states the main result of this section and is an easy consequence of the previous propositions.

**Theorem 3.21.** *For  $n \geq 0$  and a class of initial patterns  $\mathcal{I}$  with  $\mathbb{P}_0^{\mathcal{I}} \trianglelefteq \mathbb{P}_1^{\mathcal{I}}$  it holds that*

$$\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \cup \text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}}) \cap \text{co}\mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}}).$$

*Proof.* From Proposition 3.18 we obtain  $\mathbb{P}_n^{\mathcal{I}} \trianglelefteq \mathbb{P}_{n+1}^{\mathcal{I}}$  for  $n \geq 0$ . Now Proposition 3.19 implies  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}})$  for  $n \geq 0$ . From this we conclude  $\text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{co}\mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}})$  for  $n \geq 0$ . Together with Proposition 3.20 this proves the theorem.  $\square$

### 3.5 Pattern Iteration remains Starfree

In this section we show that the pattern iterator IT can be considered as a starfree iterator. Let  $\mathcal{I}$  be an arbitrary class of initial patterns and recall that SF denotes the class of starfree languages. In Theorem 3.27 we show that for  $n \geq 1$  it holds that  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{SF}$  if and only if  $\bigcup_{i \geq 0} \mathcal{FP}(\mathbb{P}_i^{\mathcal{I}}) \subseteq \text{SF}$ . For the proof of this theorem we need some auxiliary results on periodic, infinite words (see Lemma 3.22) and two modifications of the characterization of starfree languages by permutationfree DFAs (see Lemmas 3.23 and 3.25). We also make a remark on the restriction  $n \geq 1$  in Theorem 3.27.

In order to consider infinite words, we have to take over certain notions from finite words. If  $w \in A^+$  with  $w = a_1 \cdots a_m$  for alphabet letters  $a_i$ , then  $w^\infty$  denotes the infinite word  $a_1 \cdots a_m a_1 \cdots a_m \cdots$ . For  $m \geq 0$  and  $n \geq 1$  we use  $(m \bmod n)$  as an abbreviation for  $m - n \lfloor m/n \rfloor$ .

Periodic, infinite words are infinite words that can be written as  $uw^\infty$  with finite words  $u$  and  $w$ . In general there are different descriptions for one periodic, infinite word, i.e.,  $u_1 w_1^\infty = u_2 w_2^\infty$  for  $u_1 \neq u_2$  or  $w_1 \neq w_2$ . We show that if we choose a shortest  $w_1$  then the length of  $w_1$  divides the length of  $w_2$ . Note that for this purpose it suffices to show the following lemma which assumes  $u_1 = \varepsilon$ .

**Lemma 3.22.** *Let  $v \in A^+$  such that  $v^\infty \neq u'v'^\infty$  for all  $u', v' \in A^*$  with  $|v'| < |v|$ . If  $v^\infty = uw^\infty$  for some  $u, w \in A^*$  then  $|w|$  is a multiple of  $|v|$ .*

*Proof.* Let  $v$  be as above such that  $v = a_1 \cdots a_m$  for  $a_i \in A$ . We assume that  $v^\infty = uw^\infty$  for some  $u, w \in A^*$  where  $n =_{\text{def}} |w|$  is not a multiple of  $m$ . This will lead to a contradiction. First of all we define the following suffixes of  $v$  (cf. Figure 3.6).

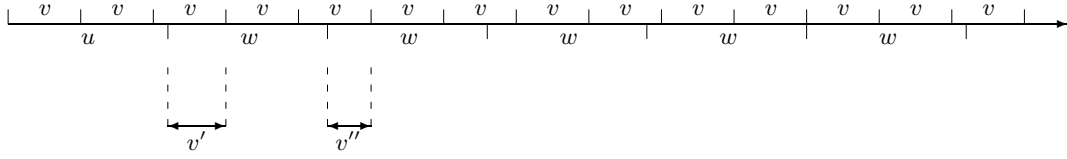
$$\begin{aligned} v' &=_{\text{def}} a_{(|u| \bmod m)+1} a_{(|u| \bmod m)+2} \cdots a_m \\ v'' &=_{\text{def}} a_{(|uw| \bmod m)+1} a_{(|uw| \bmod m)+2} \cdots a_m \end{aligned}$$

Both words  $v'$  and  $v''$  are nonempty words of length  $\leq m$ .

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

Note that  $uv'v^\infty = uv''v^\infty = v^\infty = uw^\infty$ . It follows that  $w^\infty = v'v^\infty = v''v^\infty$ . Without loss of generality we assume that  $|v'| > |v''|$ . By Definition,  $v'$  and  $v''$  are suffixes





**Fig. 3.6.** Decomposition of  $v^\infty = uw^\infty$

of  $v$ . So 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  $u, \tilde{v} \in A^*$  with  $|\tilde{v}| < |v|$  and  $v^\infty = u\tilde{v}^\infty$ . This is a contradiction to our assumption.  $\square$

The following lemma is the first extension of Theorem 1.2, a second one is stated in Lemma 3.25. By Theorem 1.2, we find a permutation (induced by some word  $w$ ) in every minimal DFA  $\mathcal{M}$  if  $L(\mathcal{M})$  is not starfree. Here we show that this permutation can even 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 not trivial, since we have to prove that the existence of such words  $z, v$  indeed induces a permutation of *distinct* states.

**Lemma 3.23.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a minimal DFA.*

$L(\mathcal{M}) \notin \text{SF} \iff$  *there exist  $w \in A^+, l \geq 2$  and distinct states  $q_1, \dots, q_l \in S$  such that*

- (i)  $w^\infty \neq zv^\infty$  for all  $z, v \in A^*$  with  $|v| < |w|$
- (ii)  $q_1 \xrightarrow{w} q_2 \xrightarrow{w} \dots \xrightarrow{w} q_l \xrightarrow{w} q_1$

*Proof.* “ $\Leftarrow$ ”: This direction is an immediate consequence of Theorem 1.2.

“ $\Rightarrow$ ”: Suppose that  $L(\mathcal{M})$  is not starfree. Using Theorem 1.2 we choose a shortest  $w \in A^+$ , some  $l \geq 2$  and distinct states  $q_1, \dots, q_l \in S$  such that  $q_1 \xrightarrow{w} q_2 \xrightarrow{w} \dots \xrightarrow{w} q_l \xrightarrow{w} q_1$ . We will show that if condition (i) is not satisfied then the choice of  $w$  was not minimal which is a contradiction.

So assume that condition (i) does not hold and choose a shortest word  $v \in A^+$  such that  $|v| < |w|$  and  $w^\infty = zv^\infty$  for some  $z \in A^*$ . Hence  $v^\infty \neq u'v'^\infty$  for all  $u', v' \in A^*$  with  $|v'| < |v|$ . 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  $w^\infty$ ). So we can apply Lemma 3.22 and obtain that  $|w|$  is a multiple of  $|v|$ . This means that  $|w| = n \cdot |v|$  and  $w = v_1v^{n-1}v_2$  for suitable  $n \geq 2$  and  $v_1, v_2 \in A^*$  with  $v = v_2v_1$ .

Now we consider the following sequence of states  $(r_i)_{i \geq 1}$ .

$$q_1 \xrightarrow{v_1} r_1 \xrightarrow{v} r_2 \xrightarrow{v} r_3 \xrightarrow{v} \dots$$

From  $w = v_1v^{n-1}v_2$  and  $v = v_2v_1$  it follows that

$$r_1 \xrightarrow{v} r_2 \xrightarrow{v} r_3 \xrightarrow{v} \dots \xrightarrow{v} r_{l \cdot n} \xrightarrow{v} r_1 \quad \text{and} \quad r_{l \cdot n} \xrightarrow{v_2} q_1.$$

Suppose that there is some  $i \geq 1$  with  $r_i = r_{i+1}$ . It follows that  $r_i \xrightarrow{v} r_i$  and  $r_i \xrightarrow{v^2} q_1$ . This implies  $q_2 = \delta(q_1, w) = \delta(r_i, v_2 w) = \delta(r_i, v^n v_2) = q_1$ , which is a contradiction to our assumption. So it follows that  $r_i \neq r_{i+1}$  for all  $i \geq 1$ .

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  $r_i, r_{i+1}, \dots, r_{j-1}$  such that

$$r_i \xrightarrow{v} r_{i+1} \xrightarrow{v} \dots \xrightarrow{v} r_{j-1} \xrightarrow{v} r_j = r_i.$$

Since  $|v| < |w|$  this is a contradiction to the choice of the shortest  $w \in A^+$  at the beginning of this proof. So we conclude that  $w^\infty \neq zv^\infty$  for all  $z, v \in A^*$  with  $|v| < |w|$ .  $\square$

Suppose we are given a DFA accepting a language that is not starfree. By Lemma 3.23 this implies that this DFA has a permutation  $q_1 \xrightarrow{w} q_2 \xrightarrow{w} \dots \xrightarrow{w} q_l \xrightarrow{w} q_1$  for some  $l \geq 2$  where  $w$  is minimal in the sense of this lemma (i.e.,  $w^\infty \neq uv^\infty$  for all  $u, v \in A^*$  with  $|v| < |w|$ ). Moreover, let  $\mathcal{I}$  be a class of initial patterns and  $n \geq 1$  such that  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}})$  is starfree. In the following lemma we show that under these assumptions we find some pattern  $p \in \mathbb{P}_n^{\mathcal{I}}$  such that for all  $r \geq 3$  the described permutation can be interpreted as a permutation induced by the bridge-word of  $\pi(p, r)$ . Furthermore, the pattern does not depend on the DFA but only on the word  $w$  and on the integer  $l \geq 2$ . The first statement of this lemma says that both permutations, the one induced by  $w$  and the one induced by the bridge-word of  $\pi(p, r)$ , take the same path through the DFA. From the second statement we obtain that the latter permutation is not trivial, i.e., it is a permutation of length  $\geq 2$ .

**Lemma 3.24.** *Let  $\mathcal{I}$  be a class of initial patterns,  $n \geq 1$ ,  $l \geq 2$ ,  $w \in A^+$  such that  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{SF}$  and  $w^\infty \neq uv^\infty$  for all  $u, v \in A^*$  with  $|v| < |w|$ . Then there exists a  $p \in \mathbb{P}_n^{\mathcal{I}}$  such that for all  $r \geq 3$  the following holds.*

1.  $w^\infty = u \cdot \left( \overline{\pi(p, r)} \right)^\infty$  for some  $u \in A^*$ .
2. The length of  $\overline{\pi(p, r)}$  is a multiple of  $|w|$ , but it is not a multiple of  $l \cdot |w|$ .

*Proof.* In order to find the pattern  $p$  we will construct a minimal DFA  $\mathcal{M}$  that accepts a language which is not starfree. This implies that  $\mathcal{M}$  has a pattern from  $\mathbb{P}_n^{\mathcal{I}}$ . This pattern will help to find the announced pattern  $p$ .

Let  $w = a_1 \dots a_m$  for letters  $a_i \in A$ , let  $S =_{\text{def}} \{s_{i,j} \mid 1 \leq i \leq l \text{ and } 1 \leq j \leq m\} \cup \{\tilde{s}\}$  be a set of states, and let  $S' =_{\text{def}} S \setminus \{s_{1,1}\}$  be a subset of accepting states. Then we define the DFA  $\mathcal{M} =_{\text{def}} (A, S, \delta, s_{1,1}, S')$  such that

$$\begin{array}{ccccccc} s_{1,1} & \xrightarrow{a_1} & s_{1,2} & \xrightarrow{a_2} & s_{1,3} & \xrightarrow{a_3} & \dots & \xrightarrow{a_{m-1}} & s_{1,m} & \xrightarrow{a_m} & s_{2,1} \\ s_{2,1} & \xrightarrow{a_1} & s_{2,2} & \xrightarrow{a_2} & s_{2,3} & \xrightarrow{a_3} & \dots & \xrightarrow{a_{m-1}} & s_{2,m} & \xrightarrow{a_m} & s_{3,1} \\ & & & & & & & & & & \vdots \\ s_{l,1} & \xrightarrow{a_1} & s_{l,2} & \xrightarrow{a_2} & s_{l,3} & \xrightarrow{a_3} & \dots & \xrightarrow{a_{m-1}} & s_{l,m} & \xrightarrow{a_m} & s_{1,1} \end{array}$$

and all remaining transitions lead to the sink  $\tilde{s}$ . First of all let us determine the language accepted by  $\mathcal{M}$ . 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  $\mathcal{M}$ . Thus a word  $v \in A^+$  is not in  $L(\mathcal{M})$  if and only if  $s_{1,1} \xrightarrow{v} s_{1,1}$ . Moreover, the only possible path which starts at  $s_{1,1}$  and which does not lead to the sink  $\tilde{s}$  looks as follows:

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

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  $L(\mathcal{M})$  if and only if it is of the form  $w^{il}$  which shows  $L(\mathcal{M}) = A^+ \setminus \{w^{il} \mid i \geq 1\}$ .

Moreover,  $\mathcal{M}$  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  $s_2 = s_{i',j'}$ . Let  $w_1$  (respectively,  $w_2$ ) be a shortest nonempty word such that  $s_1 \xrightarrow{w_1} -$  (respectively,  $s_2 \xrightarrow{w_2} -$ ). It follows that  $s_1 \xrightarrow{w_1} s_{1,1}$  and  $s_2 \xrightarrow{w_2} s_{1,1}$ , and  $w_1, w_2$  are the shortest such words. Observe that  $|w_1| = (m+1-j) + (l-i) \cdot |w|$  and  $|w_2| = (m+1-j') + (l-i') \cdot |w|$ . Furthermore, by assumption we have  $|w_1| = |w_2|$ . 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  $\mathcal{M}$  is a minimal DFA.

So we have a minimal DFA  $\mathcal{M}$  and a  $w \in A^+$  with  $s_{1,1} \xrightarrow{w} s_{2,1} \xrightarrow{w} \cdots \xrightarrow{w} s_{l,1} \xrightarrow{w} s_{1,1}$  for different states  $s_{1,1}, s_{2,1}, \dots, s_{l,1}$ . From Theorem 1.2 it follows that  $L(\mathcal{M})$  is not starfree. Moreover, from  $L(\mathcal{M}) \subseteq A^+$  and  $\mathcal{FP}(\mathbb{P}_n^Z) \subseteq \text{SF}$  we obtain that  $\mathcal{M}$  has a pattern from  $\mathbb{P}_n^Z$ . By definition there exist states  $s_1, s_2 \in S$ , a word  $z \in A^*$  and some  $p \in \mathbb{P}_n^Z$  such that

$$s_{1,1} \xrightarrow{p} s_1, \quad s_1 \xrightarrow{p} s_2, \quad s_1 \xrightarrow{z} + \quad \text{and} \quad s_1 \xrightarrow{z} -.$$

Note that  $s_1$  and  $s_2$  are different from  $\tilde{s}$ , since rejecting states are reachable from both states (e.g., the state  $\delta(s_1, \bar{p}z) = \delta(s_2, z)$ ). So we have  $s_1 = s_{i_1, j_1}$  and  $s_2 = s_{i_2, j_2}$  for suitable  $i_1, i_2, j_1, j_2$ .

By the construction of  $\mathcal{M}$  the following holds for any state  $s_{i,j} \in S$  and all  $v \in A^*$ .

$$\delta(s_{i,j}, v) \neq \tilde{s} \text{ if and only if } v \text{ is a prefix of } w_{i,j} =_{\text{def}} a_j a_{j+1} \cdots a_m \cdot w^\infty. \quad (3.4)$$

Now let  $r \geq 3$ . By Lemma 3.12, we have  $s_1 \xrightarrow{\pi(p,r)} s_2$  and  $\pi(p, r) \in \mathbb{P}_n^Z$ . Moreover, by Proposition 3.7 we have  $\overline{\pi(p, r)}, \overline{\pi(p, r)}^\circ \in A^+$  (here we need  $n \geq 1$ ) and

$$s_1 \xrightarrow{\overline{\pi(p, r)}} s_2, \quad s_1 \xrightarrow{\overline{\pi(p, r)}^\circ} s_1, \quad s_2 \xrightarrow{\overline{\pi(p, r)}^\circ} s_2.$$

From (3.4) it follows that  $\overline{\pi(p, r)}$  is a prefix of  $w_{i_1, j_1}$  and

$$w_{i_1, j_1} = \left( \overline{\pi(p, r)}^\circ \right)^\infty = w_{i_2, j_2}.$$

Since  $\delta(s_1, \overline{\pi(p, r)}) = s_2$  we have  $w_{i_1, j_1} = \overline{\pi(p, r)} \cdot w_{i_2, j_2}$ . This yields  $w_{i_1, j_1} = \overline{\pi(p, r)} \cdot w_{i_1, j_1}$ , and we obtain

$$w_{i_1, j_1} = \left( \overline{\pi(p, r)} \right)^\infty.$$

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  $w^\infty = u \cdot \left( \overline{\pi(p, r)} \right)^\infty$  which proves the first statement of this lemma.

From our assumption and Lemma 3.22 it follows that the length of  $\overline{\pi(p, r)}$  is a multiple of  $|w|$ . Suppose now that the length of  $\overline{\pi(p, r)}$  is even a multiple of  $l \cdot |w|$ . Then from the construction of  $\mathcal{M}$  it follows that  $s_2 = \delta(s_1, \overline{\pi(p, r)}) = s_1$ . This is a contradiction to the choice of  $s_1, s_2$  and  $z$  such that  $s_1 \xrightarrow{z} +$  and  $s_2 \xrightarrow{z} -$ . This shows the second statement of the lemma.  $\square$

In Theorem 1.2 starfree languages are characterized by permutationfree DFAs where a permutation is a sequence of states such that for some word  $w$  we have  $r_1 \xrightarrow{w} r_2 \xrightarrow{w} \cdots \xrightarrow{w} r_l \xrightarrow{w} r_1$ . In Lemma 3.25 below we show an extension of this theorem: One can restrict to permutations of the form  $r_1 \xrightarrow{p} r_2 \xrightarrow{p} \cdots \xrightarrow{p} r_l \xrightarrow{p} r_1$  where  $p$  is an element of some pattern class that characterizes a class of starfree languages.

**Lemma 3.25.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a minimal DFA,  $\mathcal{I}$  a class of initial patterns and  $n \geq 1$  with  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{SF}$ . Then  $L(\mathcal{M})$  is not starfree if and only if there exist a  $p \in \mathbb{P}_n^{\mathcal{I}}$ , some  $l \geq 2$  and distinct states  $r_1, r_2, \dots, r_l \in S$  such that  $r_1 \xrightarrow{p} r_2 \xrightarrow{p} \cdots \xrightarrow{p} r_l \xrightarrow{p} r_1$ .*

*Proof.* The if-part follows from Theorem 1.2 and Proposition 3.7.2.

For the only-if-part let  $r =_{\text{def}} |\mathcal{M}|$  (if  $|\mathcal{M}| < 3$  then set  $r =_{\text{def}} 3$ ) and assume that  $L(\mathcal{M})$  is not starfree. By Lemma 3.23, there exist a  $w \in A^+$ , some  $l \geq 2$  and distinct states  $q_1, q_2, \dots, q_l \in S$  such that (i)  $w^\infty \neq zv^\infty$  for all  $z, v \in A^*$  with  $|v| < |w|$  and (ii)  $q_1 \xrightarrow{w} q_2 \xrightarrow{w} \cdots \xrightarrow{w} q_l \xrightarrow{w} q_1$ . By Lemma 3.24, there exists a  $p \in \mathbb{P}_n^{\mathcal{I}}$  such that for  $p' =_{\text{def}} \pi(p, r)$  it holds that (i)  $w^\infty = u \cdot (\overline{p'})^\infty$  for some  $u \in A^*$  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  $m \geq 1$ ,  $w_1, w_2 \in A^*$  with  $w = w_2 w_1$  (note that  $m > 0$  by Proposition 3.7.1).

Now we consider the following sequence of states  $(r_i)_{i \geq 1}$ .

$$q_1 \xrightarrow{w_2} r_1 \xrightarrow{\overline{p'}} r_2 \xrightarrow{\overline{p'}} r_3 \xrightarrow{\overline{p'}} \cdots$$

Suppose there is some  $i \geq 1$  with  $r_i = r_{i+1}$ , then for a suitable  $j'$  it holds that

$$\delta(r_i, w_1) = \delta(q_1, w_2 \cdot (\overline{p'})^{i-1} w_1) = \delta(q_1, w^{m \cdot (i-1) + 1}) = q_{j'}.$$

Thus we obtain

$$\delta(q_{j'}, w^m) = \delta(r_i, w_1 w^{m-1} w_2 w_1) = \delta(r_i, \overline{p'} w_1) = \delta(r_{i+1}, w_1) = \delta(r_i, w_1) = q_{j'}.$$

Since the states  $q_1, \dots, q_l$  are pairwise different 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  $p$  and we conclude that  $r_i \neq r_{i+1}$  for all  $i \geq 1$ .

From the sequence  $(r_i)_{i \geq 1}$  we choose an earliest  $r_j$  such that there is some  $r_i = r_j$  with  $i < j$ . We have already seen that  $j - i \geq 2$ . Thus we have found a list of  $j - i \geq 2$  distinct states  $r_{i+1}, r_{i+2}, \dots, r_j$  such that

$$r_{i+1} \xrightarrow{\overline{p'}} r_{i+2} \xrightarrow{\overline{p'}} r_{i+3} \xrightarrow{\overline{p'}} \cdots \xrightarrow{\overline{p'}} r_j = r_{i+1}.$$

From Lemma 3.12.4 it follows that

$$r_{i+1} \xrightarrow{p'} r_{i+2} \xrightarrow{p'} r_{i+3} \xrightarrow{p'} \cdots \xrightarrow{p'} r_j = r_{i+1}.$$

This proves the only-if-part of the lemma.  $\square$

The iteration rule IT on the pattern side preserves the starfreeness of the characterized language classes.

**Lemma 3.26.** *If  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{SF}$  for a class of initial patterns  $\mathcal{I}$  and  $n \geq 1$ , then  $\mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}}) \subseteq \text{SF}$ .*

*Proof.* Let  $\mathcal{I}$  be a class of initial patterns and  $n \geq 1$  such that  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{SF}$ . Moreover, let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a minimal DFA such that  $L(\mathcal{M}) \subseteq A^+$  is not starfree. We will show that  $L(\mathcal{M}) \notin \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}})$ .

By Lemma 3.25, there exist a  $p_0 \in \mathbb{P}_n^{\mathcal{I}}$ , some  $l \geq 2$  and distinct states  $r_1, r_2, \dots, r_l \in S$  such that  $r_1 \xrightarrow{p_0} r_2 \xrightarrow{p_0} \dots \xrightarrow{p_0} r_l \xrightarrow{p_0} r_1$ . Now define  $p =_{\text{def}} (\overline{p_0}^{l-1}, p_0)$  and observe that  $p \in \mathbb{P}_{n+1}^{\mathcal{I}}$ . First of all we show that  $r_1 \xrightarrow{p} r_l$ . From Proposition 3.7.2 it follows that

$$r_1 \xrightarrow{\overline{p_0}} r_2 \xrightarrow{\overline{p_0}} \dots \xrightarrow{\overline{p_0}} r_l \xrightarrow{\overline{p_0}} r_1.$$

Hence  $r_1 \xrightarrow{\overline{p_0}^{l-1}} r_l$  and  $r_l \xrightarrow{p_0} r_1$ . This shows that  $p$  appears at  $r_1$ , and analogously we obtain that  $p$  appears at  $r_l$ . Moreover, from Proposition 3.7.2 it follows that  $p_0$  appears at state  $r_l$ . Together with  $\delta(r_1, \overline{p_0}^{l-1}) = r_l$  this shows  $r_1 \xrightarrow{p} r_l$ . Analogously we obtain  $r_{i+1} \xrightarrow{p} r_i$  for  $1 \leq i \leq l-1$ .

Since  $\mathcal{M}$  is minimal, there exist  $i, j$  with  $1 \leq i < j \leq l$  and a word  $z \in A^*$  such that  $\delta(r_i, z) \in S' \iff \delta(r_j, z) \notin S'$ . Since the states  $r_1, \dots, r_l$  form a cycle, there exist also an  $i$  with  $1 \leq i \leq l$  such that  $r_{i+1} \xrightarrow{z} +$  and  $r_i \xrightarrow{z} -$  (we define  $r_{l+1} =_{\text{def}} r_1$ ). Furthermore, there exists an  $x \in A^*$  with  $\delta(s_0, x) = r_{i+1}$ . Since we have already seen that  $r_{i+1} \xrightarrow{p} r_i$ , it follows that  $\mathcal{M}$  has a pattern from  $\mathbb{P}_{n+1}^{\mathcal{I}}$ . This shows  $L(\mathcal{M}) \notin \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}})$ , and it follows that  $\mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{I}}) \subseteq \text{SF}$ .  $\square$

We state the main result of this section.

**Theorem 3.27.** *For a class of initial patterns  $\mathcal{I}$  and  $n \geq 1$  the following holds.*

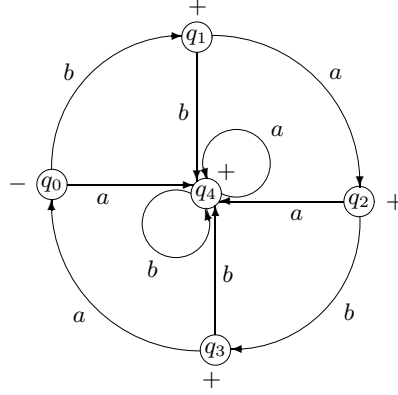
$$\bigcup_{i \geq 0} \mathcal{FP}(\mathbb{P}_i^{\mathcal{I}}) \subseteq \text{SF} \iff \mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{SF}$$

*Proof.* Let  $\mathcal{I}$  be a class of initial patterns and  $n \geq 1$ . It suffices to show that  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{SF}$  implies  $\mathcal{FP}(\mathbb{P}_i^{\mathcal{I}}) \subseteq \text{SF}$  for all  $i \geq 0$ . By Lemma 3.26, this implication holds for all  $i \geq n$ .

If  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{SF}$  then also  $\text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{SF}$ , since SF is closed under complementation. From Proposition 3.20.2 it follows that  $\mathcal{FP}(\mathbb{P}_{n-1}^{\mathcal{I}}) \subseteq \text{SF}$ . If we use this argument repeatedly, we obtain  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}}) \subseteq \text{SF} \implies \mathcal{FP}(\mathbb{P}_i^{\mathcal{I}}) \subseteq \text{SF}$  for all  $0 \leq i < n$ .  $\square$

**Remark 3.28.** It is necessary that we assume  $n \geq 1$  in Theorem 3.27. In fact, this theorem does not hold for  $n = 0$ . To see this we consider a two letter alphabet  $A = \{a, b\}$  and the class of initial patterns  $\mathcal{I} =_{\text{def}} \{(\varepsilon, \varepsilon), (\varepsilon, a), (\varepsilon, b)\}$ . With the help of the known forbidden-pattern characterization for level 1/2 of the STH (cf. Figure 3.3) we observe that  $\mathcal{FP}(\mathbb{P}_0^{\mathcal{I}}) = \mathcal{L}_{1/2} \subseteq \text{SF}$ . In contrast, we will see that  $\mathcal{FP}(\mathbb{P}_1^{\mathcal{I}}) \not\subseteq \text{SF}$ . To see this we have a look at the DFA  $\mathcal{M} = (A, S, \delta, q_0, S')$  in Figure 3.7.

It is easy to see that  $\mathcal{M}$  is a minimal DFA, and that it is not permutationfree. By Theorem 1.2, this implies  $L(\mathcal{M}) \notin \text{SF}$ . It remains to show that  $L(\mathcal{M}) \in \mathcal{FP}(\mathbb{P}_1^{\mathcal{I}})$ , i.e., we have to see that  $\mathcal{M}$  does not have a pattern from  $\mathbb{P}_1^{\mathcal{I}}$ .



**Fig. 3.7.** A DFA  $\mathcal{M}$  with  $L(\mathcal{M}) \in \mathcal{FP}(\mathbb{P}_1^{\mathcal{I}}) \setminus \text{SF}$  for  $\mathcal{I} = \{(\varepsilon, \varepsilon), (\varepsilon, a), (\varepsilon, b)\}$ .

Assume that  $\mathcal{M}$  has a pattern from  $\mathbb{P}_1^{\mathcal{I}}$ , i.e., there are states  $q', q''$ , words  $x, z$  and a  $p \in \mathbb{P}_1^{\mathcal{I}}$  such that

$$q_0 \xrightarrow{x} q', \quad q' \xrightarrow{p} q'', \quad q' \xrightarrow{z} + \quad \text{and} \quad q'' \xrightarrow{z} -.$$

Choose suitable  $w_i \in A^+$  and  $p_i \in \mathcal{I}$  such that  $p = (w_0, p_0, \dots, w_m, p_m)$ . So we have  $\bar{p} = w_0 \cdots w_m$  and  $\bar{p}^\circ = w_0 \bar{p}_0 \cdots w_m \bar{p}_m$ . Note that  $q', q'' \in \{q_0, q_1, q_2, q_3\}$ , since rejecting states are reachable from  $q'$  and  $q''$ . It follows that  $\bar{p}$  and  $\bar{p}^\circ$  are alternating sequences of letters  $a, b$ , since all other words lead to the sink  $q_4$ .

Note that  $\bar{p}_i \in \{\varepsilon, a, b\}$  for all  $0 \leq i \leq m$ , and assume that  $\bar{p}_j \neq \varepsilon$  for some  $0 \leq j < m$ . It follows that either  $\bar{p}_j$  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  $\bar{p}$  is an alternating sequences of letters  $a, b$ . It follows that  $\bar{p}^\circ = \bar{p} \cdot \bar{p}_m$  where  $\bar{p}_m \in \{a, b\}$ , since  $\bar{p} \neq \bar{p}^\circ$ . Without loss of generality we assume that  $\bar{p}_m = a$ . Hence  $\delta(q'', a) = \delta(q', \bar{p}a) = q'$ , and it follows that  $\delta(q', a) = q_4$  and  $\delta(q'', b) = q_4$ . Therefore, at least one of the states  $q'$  and  $q''$  does not have a  $\bar{p}^\circ$ -loop. This is a contradiction to our assumption, and it follows that  $\mathcal{M}$  does not have a pattern from  $\mathbb{P}_1^{\mathcal{I}}$ . This shows  $\mathcal{FP}(\mathbb{P}_1^{\mathcal{I}}) \not\subseteq \text{SF}$ .

Note that if some DFA has a pattern from  $\mathbb{P}_1^{\mathcal{I}}$  then there are states  $s_1, s_2$  and words  $w, z$  such that  $s_0 \xrightarrow{w} s_1 \xrightarrow{w} s_2$ ,  $s_1 \xrightarrow{z} +$  and  $s_2 \xrightarrow{z} -$ . Hence we find states  $s'_1, s'_2$  on the path  $s_1 \xrightarrow{w} s_2$  and a letter  $c \in A$  such that  $s'_1 \xrightarrow{c} s'_2$ ,  $s'_1 \xrightarrow{z} +$  and  $s'_2 \xrightarrow{z} -$ . This shows that  $\mathcal{FP}(\mathbb{P}_0^{\mathcal{I}}) \subseteq \mathcal{FP}(\mathbb{P}_1^{\mathcal{I}})$ . So we even obtain  $\mathcal{FP}(\mathbb{P}_0^{\mathcal{I}}) \subseteq \text{SF}$  and  $\mathcal{FP}(\mathbb{P}_0^{\mathcal{I}}) \subseteq \mathcal{FP}(\mathbb{P}_1^{\mathcal{I}}) \not\subseteq \text{SF}$ .

### 3.6 Decidability of the Forbidden-Pattern Classes

In this section we treat the decidability aspects of the forbidden-pattern classes. It will turn out that  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}})$  is decidable in nondeterministic logarithmic space whenever a certain

family of decision problems for the class of initial patterns  $\mathcal{I}$  (see Definition 3.30 for  $n = 0$ ) is decidable in these space bounds. This family consists of the following decision problems for each constant  $k \geq 1$ : Decide on input of some DFA  $\mathcal{M}$ ,  $k$  states of  $\mathcal{M}$  and  $k$  pairs of states of  $\mathcal{M}$  whether there is some  $p \in \mathcal{I}$  appearing at each of the given single states and connecting each of the given pairs. Note that the decidability of the forbidden-pattern classes has to depend on the class of initial patterns, since an undecidable set  $\mathcal{I}$  (which can be easily constructed) leads to undecidable forbidden-pattern classes.

We start with the definition of two problems addressing the question of the existence of paths and patterns that appear simultaneously in a given DFA. In the Lemmas 3.32 and 3.33 we investigate the decidability of these problems, and at the end this leads to a decidability result for the forbidden-pattern classes (see Theorem 3.35).

**Definition 3.29.** *Let  $k \geq 1$ . We define  $\text{REACH}_k$  to be the set of pairs  $(\mathcal{M}, W)$  such that:*

1.  $\mathcal{M} = (A, S, \delta, s_0, S')$  is a DFA
2.  $W \subseteq S \times S$  with  $|W| \leq k$
3. There exists a word  $w \in A^+$  such that in  $\mathcal{M}$  we have  $s \xrightarrow{w} t$  for all  $(s, t) \in W$ .

**Definition 3.30.** *Let  $\mathcal{I}$  be a class of initial patterns,  $n \geq 0$  and  $k \geq 1$ . We define  $\text{PATTERN}_{n,k}^{\mathcal{I}}$  to be the set of all triples  $(\mathcal{M}, T_1, T_2)$  such that the following holds:*

1.  $\mathcal{M} = (A, S, \delta, s_0, S')$  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  $p \in \mathbb{P}_n^{\mathcal{I}}$  such that  $p$  appears at  $s$  and  $q \xrightarrow{p} r$  for all  $s \in T_1$  and  $(q, r) \in T_2$ .

**Remark 3.31.** 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  $\mathcal{M} = (A, S, \delta, s_0, S')$ . For this we fix arbitrary orderings on the sets  $A$ ,  $S$  and  $S'$ , such that we obtain  $A = \{a_1, \dots, a_{|A|}\}$ ,  $S = \{s_1, \dots, s_{|S|}\}$  (one of them is the starting state  $s_0$ ) and  $S' = \{s'_1, \dots, s'_{|S'|}\}$ . Moreover, we identify elements  $a_i \in A$  and  $s_j \in S$  with their index numbers  $i$  and  $j$ . Now we can encode  $\mathcal{M}$  in the following way.

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

The sets  $W$ ,  $S_1$ ,  $S_2$  in the Definitions 3.29 and 3.30 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  $(\mathcal{M}, W)$  and  $(\mathcal{M}, S_1, S_2)$ . 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  $(\mathcal{M}, W)$  (respectively, of a triple  $(\mathcal{M}, S_1, S_2)$ ). Therefore, in forthcoming investigation of algorithms we may assume that all inputs are valid representations.

**Lemma 3.32.** *For  $k \geq 1$  we have  $\text{REACH}_k \in \text{NL}$ .*

*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  $s_1, \dots, s_k$  and  $t_1, \dots, t_k$  (some may take the same value if  $|W| < k$ ). Now we guess a word  $w \in A^+$  letter by letter, and we simultaneously follow the paths which start at  $s_1, \dots, s_k$  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  $1 \leq i \leq k$ .  $\square$

We consider oracle machines working in nondeterministic logarithmic space which have the following access to the oracle. The machine has a (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 3.33.** *Let  $\mathcal{I}$  be a class of initial patterns. Then  $\text{PATTERN}_{n,k}^{\mathcal{I}} \in \text{NL}^{\text{PATTERN}_{(n-1),3k}^{\mathcal{I}}}$  for each  $n \geq 1$  and each  $k \geq 1$ .*

*Proof.* In Table 3.1 we describe a nondeterministic algorithm having access to a  $\text{REACH}_{4k}$  oracle and to a  $\text{PATTERN}_{(n-1),3k}^{\mathcal{I}}$  oracle. The notations in this table are adopted from the Figures 3.8 and 3.9. We will show that this algorithm works in logarithmic space and decides  $\text{PATTERN}_{n,k}^{\mathcal{I}}$ . By Lemma 3.32 we have  $\text{REACH}_k \in \text{NL}$ . Since the access to an NL oracle does not rise the power of an NL machine, i.e.,  $\text{NL}^{\text{NL}} = \text{NL}$  [RST82, Sze87, Imm88], we can go without the  $\text{REACH}_{4k}$  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  $\mathcal{M}$  is stored on the input-tape, we can actually write down the queries  $(\mathcal{M}, W)$  and  $(\mathcal{M}, T'_1, T'_2)$  without making any nondeterministic branches.

Let us analyze the space on the working-tape which is needed on input  $(\mathcal{M}, T_1, T_2)$ . 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  $\mathcal{M}$ , 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 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  $\text{PATTERN}_{n,k}^{\mathcal{I}}$ . First of all we want to see that the computation has an accepting path if  $(\mathcal{M}, T_1, T_2) \in \text{PATTERN}_{n,k}^{\mathcal{I}}$ . For this let  $p = (w_0, p_0, \dots, w_m, p_m) \in \mathbb{P}_n^{\mathcal{I}}$  be a witnessing pattern (see Definition 3.30.4). We denote the involved states of the appearance of  $p$  at  $s_i$  as in Figure 3.8, and we denote the involved states of the connection of  $q_j, r_j$  via  $p$  as in Figure 3.9. Now consider that path of the computation where we carry out exactly  $m+1$  passes of the loop



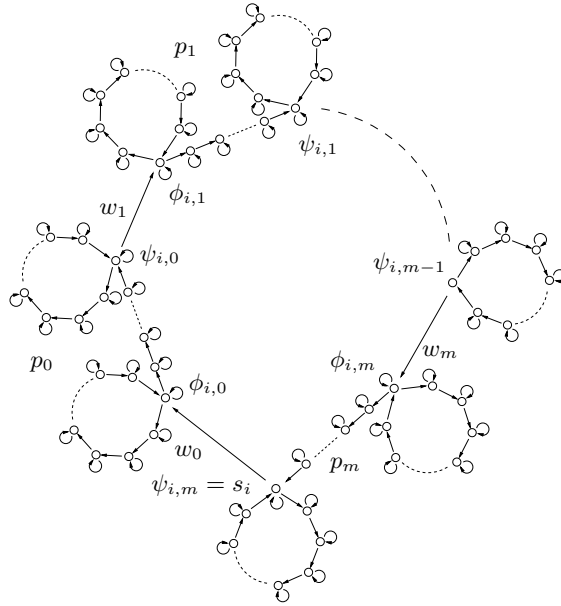


Fig. 3.8. Example for the appearance of a pattern at state  $s_i$ .

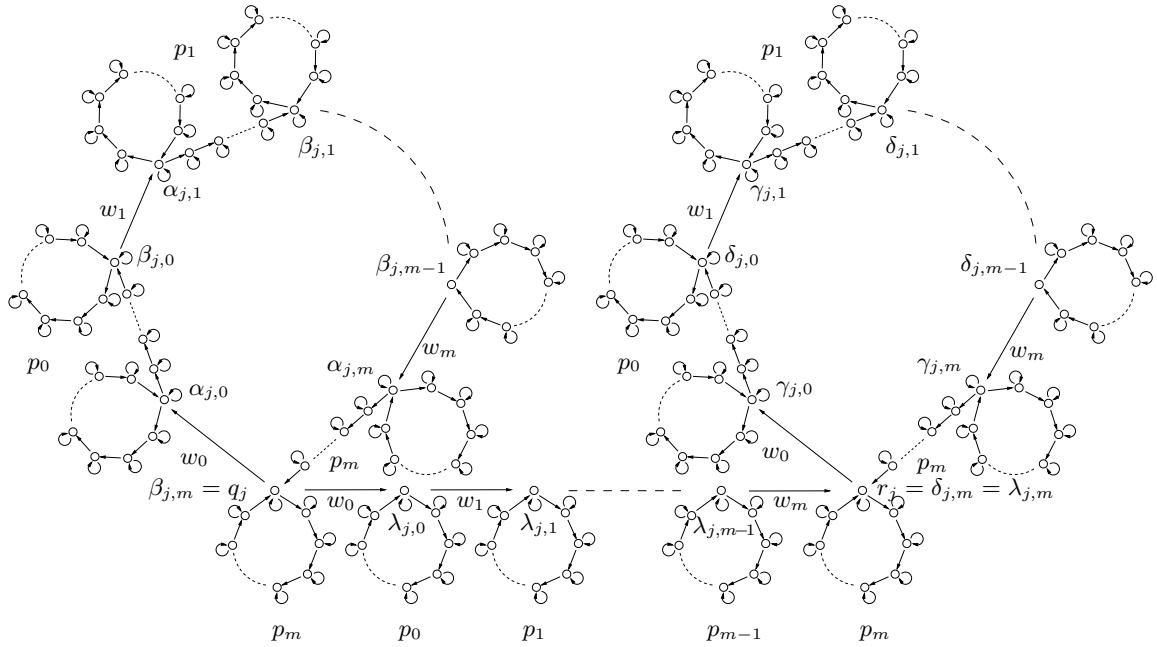


Fig. 3.9. Example for the connection of two states  $q_j, r_j$  via a pattern.

Step, Label	Command	Remark
1.	Let $t_1 :=  T_1 $ , $t_2 :=  T_2 $ and let $s_i, q_j, r_j$ such that $T_1 = \{s_1, \dots, s_{t_1}\}$ and $T_2 = \{(q_1, r_1), \dots, (q_{t_2}, r_{t_2})\}$ .	Note that $t_1, t_2$ are bounded by the constant $k$ . We have to decide whether there is a $p \in \mathbb{P}_n^Z$ such that (i) $p$ appears at all $s_i$ (Figure 3.8) and (ii) all $q_j, r_j$ are connected via $p$ (Figure 3.9).
2.	For $1 \leq i \leq t_1$ and $1 \leq j \leq t_2$ let: $\psi_i^{\text{start}} := s_i$ $\beta_j^{\text{start}} := q_j$ $\delta_j^{\text{start}} := r_j$ $\lambda_j^{\text{start}} := q_j$	Variables marked with ‘start’ contain the starting point from where we have to guess and check the next fragment of the pattern.
3. loop:	Guess states $\phi_i, \psi_i$ 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: $T_1' := \{ \lambda_j \mid 1 \leq j \leq t_2 \}$ $T_2' := \{ (\phi_i, \psi_i) \mid 1 \leq i \leq t_1 \} \cup \{ (\alpha_j, \beta_j) \mid 1 \leq j \leq t_2 \} \cup \{ (\gamma_j, \delta_j) \mid 1 \leq j \leq t_2 \}$ $W := \{ (\psi_i^{\text{start}}, \phi_i) \mid 1 \leq i \leq t_1 \} \cup \{ (\beta_j^{\text{start}}, \alpha_j) \mid 1 \leq j \leq t_2 \} \cup \{ (\delta_j^{\text{start}}, \gamma_j) \mid 1 \leq j \leq t_2 \} \cup \{ (\lambda_j^{\text{start}}, \lambda_j) \mid 1 \leq j \leq t_2 \}$	The guessed states correspond to Figures 3.8 and 3.9. 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 $\phi_{i,l}, \psi_{i,l}, \alpha_{j,l}, \beta_{j,l}, \gamma_{j,l}, \delta_{j,l}, \lambda_{j,l}$ (resp.). 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 3.8 and 3.9).
4.	Ask the following queries and reject when a negative answer is given. $(\mathcal{M}, W) \in \text{REACH}_{4k}$ $(\mathcal{M}, T_1', T_2') \in \text{PATTERN}_{(n-1), 3k}^Z$	If at least one negative answer is given then the states guessed at the previous step do not correspond to a pattern from $\mathbb{P}_n^Z$ .
5.	For $1 \leq i \leq t_1$ and $1 \leq j \leq t_2$ let: $\psi_i^{\text{start}} := \psi_i$ $\beta_j^{\text{start}} := \beta_j$ $\delta_j^{\text{start}} := \delta_j$ $\lambda_j^{\text{start}} := \lambda_j$	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 $1 \leq j \leq t_2$ : $\psi_i = s_i$ $\beta_j = q_j$ $\delta_j = r_j$ $\lambda_j = r_j$	It remains to check whether the guessed loops have reached their starting points, and whether the path which was guessed via $\lambda_i$ leads from $q_i$ to $r_i$ .

**Table 3.1.** An algorithm which decides  $(\mathcal{M}, T_1, T_2) \in \text{PATTERN}_{n,k}^Z$  on input of a DFA  $\mathcal{M} = (A, S, \delta, s_0, S')$  and sets  $T_1 \subseteq S$  and  $T_2 \subseteq S \times S$  with  $|T_1|, |T_2| \leq k$ .

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  $(\mathcal{M}, T_1, T_2)$  has accepting paths, and fix one of them. 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  $(\mathcal{M}, W) \in \text{REACH}_{4k}$  and  $(\mathcal{M}, T'_1, T'_2) \in \text{PATTERN}_{(n-1), 3k}^{\mathcal{I}}$  (otherwise the fixed path would be rejecting). It follows that for each pass  $l$  there exists a word  $w_l \in A^+$  witnessing  $(\mathcal{M}, W) \in \text{REACH}_{4k}$ , and there exists a pattern  $p_l \in \mathbb{P}_{n-1}^{\mathcal{I}}$  witnessing  $(\mathcal{M}, T'_1, T'_2) \in \text{PATTERN}_{(n-1), 3k}^{\mathcal{I}}$ . Now define  $p =_{\text{def}} (w_0, p_0, \dots, w_m, p_m)$ . 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  $p$ .  $\square$

**Corollary 3.34.** *Let  $\mathcal{I}$  be a class of initial patterns such that  $\text{PATTERN}_{0,k}^{\mathcal{I}} \in \text{NL}$  for all  $k \geq 1$ . Then  $\text{PATTERN}_{n,k}^{\mathcal{I}} \in \text{NL}$  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 3.33 and the fact that  $\text{NL}^{\text{NL}} = \text{NL}$  [RST82, Sze87, Imm88].  $\square$

Now it is easy to prove the main result of this section. It says that it can be efficiently tested whether a DFA has a pattern from  $\mathbb{P}_n^{\mathcal{I}}$  provided that efficient tests for the class of initial patterns are available. In particular, for a suitable class of initial patterns  $\mathcal{I}$  (i.e.,  $\mathcal{I}$  satisfies the assumption of our theorem) this implies the efficient decidability of all forbidden-pattern classes  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}})$ .

**Theorem 3.35.** *Let  $\mathcal{I}$  be a class of initial patterns with  $\text{PATTERN}_{0,k}^{\mathcal{I}} \in \text{NL}$  for each  $k \geq 1$ . Then for a fixed  $n \geq 0$  it is decidable in nondeterministic logarithmic space whether a given DFA has a pattern from  $\mathbb{P}_n^{\mathcal{I}}$ .*

*Proof.* On input  $\mathcal{M} = (A, S, \delta, s_0, S')$  we guess states  $s_1, s_2, s^+, s^- \in S$  and check whether  $s^+ \in S'$  and  $s^- \notin S'$ . Now we test  $(\mathcal{M}, \{(s_0, s_1)\}) \in \text{REACH}_1$  and  $(\mathcal{M}, \{(s_1, s^+), (s_2, s^-)\}) \in \text{REACH}_2$  which is possible in NL by Lemma 3.32. It remains to check whether  $(\mathcal{M}, \emptyset, \{(s_1, s_2)\}) \in \text{PATTERN}_{n,1}^{\mathcal{I}}$  which is also possible in NL by Corollary 3.34.  $\square$

### 3.7 Consequences for Concatenation Hierarchies

In this section we consider two particular classes of initial patterns  $\mathcal{L}$  and  $\mathcal{B}$ . We will see that the emerging classes of languages are closely related to the DDH and the STH. In particular we show the following for the classes  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$  and  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{B}})$ .

- Both form strict and decidable hierarchies which exhaust the class of starfree languages.
- They have the same inclusion structure as it is known from the concatenation hierarchies.
- They contain the level  $n+1/2$  and does *not* contain the level  $n+3/2$  of the corresponding concatenation hierarchy.

Moreover, in chapter 4 we show more similarities between these forbidden-pattern classes and the concatenation hierarchies, namely that they coincide on some lower levels. We start with the formal definition of the class of initial patterns  $\mathcal{L}$  and  $\mathcal{B}$ .

**Definition 3.36.** *We define the following classes of 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 classes of initial patterns.

### 3.7.1 Inclusion Relations between DDH, STH and Forbidden-Pattern Classes

First of all let us clarify the inclusion structure of the classes  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$  and  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{B}})$ . In Theorem 3.38 below we show that this is the same inclusion structure as in the case of the concatenation hierarchies DDH and STH. In order to apply Proposition 3.18 we have to show the following.

**Lemma 3.37.** *It holds that  $\mathbb{P}_0^{\mathcal{L}} \trianglelefteq \mathbb{P}_0^{\mathcal{B}}$ ,  $\mathbb{P}_0^{\mathcal{B}} \trianglelefteq \mathbb{P}_1^{\mathcal{L}}$ ,  $\mathbb{P}_0^{\mathcal{B}} \trianglelefteq \mathbb{P}_1^{\mathcal{B}}$  and  $\mathbb{P}_0^{\mathcal{L}} \trianglelefteq \mathbb{P}_1^{\mathcal{L}}$ .*

*Proof.* To see  $\mathbb{P}_0^{\mathcal{L}} \trianglelefteq \mathbb{P}_0^{\mathcal{B}}$ , let  $p_1 = (v, w) \in \mathbb{P}_0^{\mathcal{B}} = A^+ \times A^+$  and define  $p_2 = (\varepsilon, w)$ . Obviously,  $p_2 \in \mathbb{P}_0^{\mathcal{L}}$ . Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA and  $s, s_1, s_2 \in S$  such that  $p_1$  appears at  $s$  and  $s_1 \xrightarrow{p_1} s_2$ . Clearly,  $s \xrightarrow{\varepsilon} s$ , so also  $p_2$  appears at  $s$ . Moreover we have  $s_1 \xrightarrow{w} s_2$ , so  $s_1 \xrightarrow{p_2} s_2$ . This shows  $\mathbb{P}_0^{\mathcal{L}} \trianglelefteq \mathbb{P}_0^{\mathcal{B}}$ .

Now let  $p_1 = (w_0, p'_0, \dots, w_m, p'_m) \in \mathbb{P}_1^{\mathcal{L}}$  with  $w_i \in A^+$  and  $p'_i = (l_i, b_i) \in \mathbb{P}_0^{\mathcal{L}}$  for all  $0 \leq i \leq m$ . Define  $p_2 =_{\text{def}} (\overline{p_1}^\circ, \overline{p_1})$ . By Proposition 3.7.1 we have  $p_2 \in A^+ \times A^+ = \mathbb{P}_0^{\mathcal{B}}$ . Again, let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA and  $s, s_1, s_2 \in S$ . First assume that  $p_1$  appears at  $s$ . By Proposition 3.7.3 we have  $\delta(s, \overline{p_1}^\circ) = s$ , and hence also  $p_2$  appears at  $s$ . Now suppose  $s_1 \xrightarrow{p_1} s_2$ . Since  $p_1$  appears at  $s_1$  and at  $s_2$ , also  $p_2$  does so. Furthermore,  $s_1 \xrightarrow{\overline{p_1}} s_2$  by Proposition 3.7.2. So  $s_1 \xrightarrow{p_2} s_2$  which shows  $\mathbb{P}_0^{\mathcal{B}} \trianglelefteq \mathbb{P}_1^{\mathcal{L}}$ .

Analogously we prove  $\mathbb{P}_0^{\mathcal{B}} \trianglelefteq \mathbb{P}_1^{\mathcal{B}}$ . Finally, taking together  $\mathbb{P}_0^{\mathcal{L}} \trianglelefteq \mathbb{P}_0^{\mathcal{B}}$  and  $\mathbb{P}_0^{\mathcal{B}} \trianglelefteq \mathbb{P}_1^{\mathcal{L}}$  we get  $\mathbb{P}_0^{\mathcal{L}} \trianglelefteq \mathbb{P}_1^{\mathcal{L}}$ .  $\square$

**Theorem 3.38.** *For  $n \geq 0$  the following holds.*

1.  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}}) \subseteq \mathcal{FP}(\mathbb{P}_n^{\mathcal{B}})$
2.  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{B}}) \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{L}})$
3.  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}}) \cup \text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}}) \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{L}}) \cap \text{co}\mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{L}})$
4.  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{B}}) \cup \text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{B}}) \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{B}}) \cap \text{co}\mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{B}})$

*Proof.* We have  $\mathbb{P}_0^{\mathcal{L}} \trianglelefteq \mathbb{P}_0^{\mathcal{B}}$  and  $\mathbb{P}_0^{\mathcal{B}} \trianglelefteq \mathbb{P}_1^{\mathcal{L}}$  by Lemma 3.37. By Proposition 3.18 this implies  $\mathbb{P}_n^{\mathcal{L}} \trianglelefteq \mathbb{P}_n^{\mathcal{B}}$  and  $\mathbb{P}_n^{\mathcal{B}} \trianglelefteq \mathbb{P}_{n+1}^{\mathcal{L}}$  for all  $n \geq 0$ . Now Proposition 3.19 shows the statements 1 and 2. From Lemma 3.37 we also know  $\mathbb{P}_0^{\mathcal{L}} \trianglelefteq \mathbb{P}_1^{\mathcal{L}}$  and  $\mathbb{P}_0^{\mathcal{B}} \trianglelefteq \mathbb{P}_1^{\mathcal{B}}$ . Together with Theorem 3.21 we get the remaining statements.  $\square$

Note that these inclusion relations are similar to those of the concatenation hierarchies (Propositions 2.3 and 2.4). However, the following theorem shows that the connections between forbidden-pattern classes and classes of concatenation hierarchies are even closer.

**Theorem 3.39.** *For  $n \geq 0$  the following holds.*

1.  $\mathcal{L}_{n+1/2} \subseteq \mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$
2.  $\mathcal{B}_{n+1/2} \subseteq \mathcal{FP}(\mathbb{P}_n^{\mathcal{B}})$

*Proof.* First of all we show that  $\mathcal{L}_{1/2} \subseteq \mathcal{FP}(\mathbb{P}_0^{\mathcal{L}})$ . For this let  $L$  be a regular language such that  $L \notin \mathcal{FP}(\mathbb{P}_0^{\mathcal{L}})$ . We will show that  $L \notin \mathcal{L}_{1/2}$ . By definition, the minimal DFA  $\mathcal{M} = (A, S, \delta, s_0, S')$  with  $L(\mathcal{M}) = L$  has a pattern from  $\mathbb{P}_0^{\mathcal{L}}$ , i.e., there exist states  $s_1, s_2$ , words  $x, z$  and some  $p = (\varepsilon, w) \in \mathbb{P}_0^{\mathcal{L}}$  such that

$$s_0 \xrightarrow{x} s_1, \quad s_1 \xrightarrow[p]{\varepsilon} s_2, \quad s_1 \xrightarrow{z} + \quad \text{and} \quad s_2 \xrightarrow{z} -.$$

So we have  $xz \in L$ , and from Definition 3.2 it follows that  $s_1 \xrightarrow{w} s_2$  which implies  $xwz \notin L$ .

Suppose  $L \in \mathcal{L}_{1/2}$ , this will lead to a contradiction. By definition,  $L$  is a finite union of languages of the form  $A^*a_0A^*\cdots a_mA^*$  for alphabet letters  $a_i \in A$ . Since  $xz \in L$ , there exist letters  $a_0, \dots, a_m \in A$  such that  $xz \in A^*a_0A^*\cdots a_mA^* \subseteq L$ . This means that there exists an increasing sequence of positions in the word  $xz$  where we find the letters  $a_0, \dots, a_m$  in this ordering. It follows that also the word  $xwz$  has this property. Thus  $xwz \in A^*a_0A^*\cdots a_mA^* \subseteq L$  which is a contradiction, since we have already seen that  $xwz \notin L$ . This shows  $L \notin \mathcal{L}_{1/2}$  and it follows  $\mathcal{L}_{1/2} \subseteq \mathcal{FP}(\mathbb{P}_0^{\mathcal{L}})$ .

Now let us show  $\mathcal{B}_{1/2} \subseteq \mathcal{FP}(\mathbb{P}_0^{\mathcal{B}})$  where we proceed analogously to the proof above. Let  $L$  be a regular language with  $L \notin \mathcal{FP}(\mathbb{P}_0^{\mathcal{B}})$ , and let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be the minimal DFA with  $L(\mathcal{M}) = L$ . By definition,  $\mathcal{M}$  has a pattern from  $\mathbb{P}_0^{\mathcal{B}}$ , i.e., there exist states  $s_1, s_2$ , words  $x, z$  and some  $p = (v, w) \in \mathbb{P}_0^{\mathcal{B}}$  such that

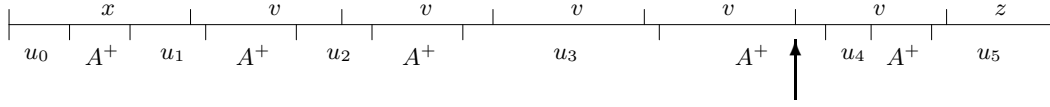
$$s_0 \xrightarrow{x} s_1, \quad s_1 \xrightarrow[p]{\varepsilon} s_2, \quad s_1 \xrightarrow{z} + \quad \text{and} \quad s_2 \xrightarrow{z} -.$$

In particular we have  $s_1 \xrightarrow{w} s_2$  and there are  $v$ -loops at the states  $s_1$  and  $s_2$ . It follows that  $xv^i z \in L$  and  $xv^i wv^j z \notin L$  for all  $i, j$ .

We assume that  $L \in \mathcal{B}_{1/2}$ , this will lead to a contradiction. By definition,  $L$  is a finite union of languages of the form  $u_0A^+u_1\cdots A^+u_m$  for words  $u_i \in A^*$ . Since all words  $xv^i z$  are elements of  $L$  there exists some sufficiently large  $r \geq 2$  such that

- $xv^r z \in u_0A^+u_1\cdots A^+u_m$  where  $u_0A^+u_1\cdots A^+u_m$  is an element of the finite union describing  $L$  and
- $xv^r z$  can be decomposed as  $xv^r z = xv^i \cdot v^j z$  such that  $xv^i \in u_0A^+u_1\cdots A^+u_l A^+$  and  $v^j z \in A^+u_{l+1}A^+u_{l+2}\cdots A^+u_l$  for a suitable  $l \geq 0$ .

This is illustrated in Figure 3.10. Here we have  $xv^5 z \in u_0A^+u_1\cdots A^+u_5$  and the arrow marks the position where we have to decompose the word  $xv^5 z$ . On the one hand at this position we find a border between two neighboring factors  $v$ , and on the other hand it is located in an area that is assigned to some  $A^+$ . It follows that at this position we can insert any word without leaving the language  $L$ . In particular this implies  $xv^i wv^j z \in L$  which is a contradiction. This shows  $L \notin \mathcal{B}_{1/2}$  and it follows  $\mathcal{B}_{1/2} \subseteq \mathcal{FP}(\mathbb{P}_0^{\mathcal{B}})$ .



**Fig. 3.10.** Decomposition of the word  $xv^r z$

Now we proceed by induction on  $n$ . So assume that we have shown the lemma for some  $n \geq 0$  and consider  $\mathcal{L}_{n+3/2}$ . The induction hypothesis says that  $\mathcal{L}_{n+1/2} \subseteq \mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$  which implies  $\text{co}\mathcal{L}_{n+1/2} \subseteq \text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$ . Together with Theorem 2.10 and the monotony of Pol we get  $\mathcal{L}_{n+3/2} = \text{Pol}(\text{co}\mathcal{L}_{n+1/2}) \subseteq \text{Pol}(\text{co}\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}}))$ . Now we apply Theorem 3.16 and obtain  $\mathcal{L}_{n+3/2} \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{L}})$ . Analogously we show  $\mathcal{B}_{n+3/2} \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^{\mathcal{B}})$ .  $\square$

### 3.7.2 The Forbidden-Pattern Classes are Starfree

The forbidden-pattern hierarchies exhaust the class of starfree languages.

**Theorem 3.40.** *It holds that  $\bigcup_{n \geq 0} \mathcal{FP}(\mathbb{P}_n^{\mathcal{B}}) = \bigcup_{n \geq 0} \mathcal{FP}(\mathbb{P}_n^{\mathcal{L}}) = \text{SF}$ .*

*Proof.* From the Theorems 2.5, 3.39 and 3.38 we get

$$\text{SF} = \bigcup_{n \geq 0} \mathcal{B}_{n+1/2} \subseteq \bigcup_{n \geq 0} \mathcal{FP}(\mathbb{P}_n^{\mathcal{B}}) = \bigcup_{n \geq 0} \mathcal{FP}(\mathbb{P}_n^{\mathcal{L}}).$$

So it remains to show that

$$\bigcup_{n \geq 0} \mathcal{FP}(\mathbb{P}_n^{\mathcal{L}}) \subseteq \text{SF}.$$

By Theorem 3.27 it suffices to show  $\mathcal{FP}(\mathbb{P}_1^{\mathcal{L}}) \subseteq \text{SF}$ . Let  $L$  be a regular language that is not starfree, we will show that  $L \notin \mathcal{FP}(\mathbb{P}_1^{\mathcal{L}})$ . We denote by  $\mathcal{M}$  the minimal DFA with  $L(\mathcal{M}) = L$ . By Theorem 1.2 we know that  $\mathcal{M}$  is not permutationfree, i.e., there exist a word  $w \in A^+$ , some  $l \geq 2$  and distinct states  $r_1, r_2, \dots, r_l \in S$  such that

$$r_1 \xrightarrow{w} r_2 \xrightarrow{w} \dots \xrightarrow{w} r_l \xrightarrow{w} r_1.$$

Since  $\mathcal{M}$  is minimal there exist a word  $z$  and states  $r_i, r_j$  from this cycle such that  $r_i \xrightarrow{z} +$  and  $r_j \xrightarrow{z} -$ . Since the states  $r_1, \dots, r_l$  form a cycle, there exists a  $k$  with  $1 \leq k \leq l$  such that  $r_k \xrightarrow{z} +$  and  $r_{k+1} \xrightarrow{z} -$  (we define  $r_{l+1} =_{\text{def}} r_1$ ). Without loss of generality we may assume that  $k = 1$ , i.e.,  $r_1 \xrightarrow{z} +$  and  $r_2 \xrightarrow{z} -$  (otherwise we rename the states).

Let  $p_0 =_{\text{def}} (\varepsilon, w^{l-1})$ ,  $p =_{\text{def}} (w, p_0)$  and observe that  $p_0 \in \mathbb{P}_0^{\mathcal{L}}$  and  $p \in \mathbb{P}_1^{\mathcal{L}}$ . We want to show  $r_1 \xrightarrow{p} r_2$ . For this observe that

$$r_1 \xrightarrow{w} r_2 \xrightarrow{p_0} r_1 \quad \text{and} \quad r_2 \xrightarrow{w} r_3 \xrightarrow{p_0} r_2.$$

It follows that  $p$  appears at  $r_1$  and at  $r_2$ . Moreover,  $r_1 \xrightarrow{w} r_2$  and  $p_0$  appears at  $r_2$  (since  $r_2 \xrightarrow{\varepsilon} r_2$ ). This shows  $r_1 \xrightarrow{p} r_2$  and it follows that  $\mathcal{M}$  has a pattern from  $\mathbb{P}_1^{\mathcal{L}}$ . This implies  $L \notin \mathcal{FP}(\mathbb{P}_1^{\mathcal{L}})$  which shows  $\mathcal{FP}(\mathbb{P}_1^{\mathcal{L}}) \subseteq \text{SF}$ .  $\square$

### 3.7.3 The Hierarchies of Forbidden-Pattern Classes are Strict

We want to show the strictness of the two hierarchies  $\{\mathcal{FP}(\mathbb{P}_n^B)\}$  and  $\{\mathcal{FP}(\mathbb{P}_n^C)\}$  in a certain way, namely we take witnessing languages from [Tho84] that were used there to separate the classes of the DDH. As remarked in [Tho84], these languages can also be used to show that the STH 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 DDH was defined exactly as we did here (namely, not taking  $\varepsilon$  into account).

We assume in this subsection that  $A = \{a, b\}$ . In Theorem 3.44 we will separate the instances of the forbidden-pattern hierarchies defined for this alphabet. In Remark 3.45 below we show that this separation can be extended to the general case.

We start with the definition of a particular family of languages of  $A^+$  from [Tho84]. Denote for  $w \in A^+$  by  $|w|_a$  (respectively,  $|w|_b$ ) the number of occurrences of  $a$  (respectively,  $b$ ) 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 (i)  $L_n \in \mathcal{B}_n$  for  $n \geq 1$  and (ii)  $L_n \notin \mathcal{B}_{n-1}$  for  $n \geq 2$  (there these languages were denoted as  $L_n^+$ ). Moreover, it is easy to see that for the DFA given in Figure 3.11 it holds that  $L_n = L(\mathcal{M}_n)$ .

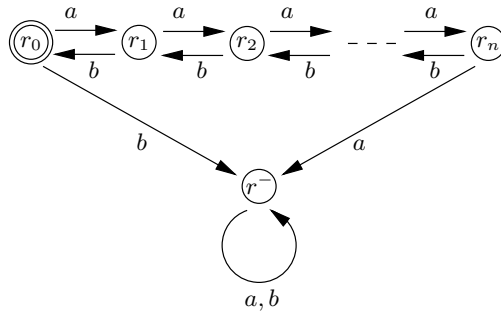


Fig. 3.11. DFA  $\mathcal{M}_n$  where  $r_n$  is the only accepting state. It holds that  $L(\mathcal{M}_n) = L_n$ .

**Lemma 3.41** ([Tho84]). *Let  $n \geq 1$ . Then  $L_n \in \mathcal{B}_n$ .*

If we consider the states  $r_0, r_1, \dots, r_n$  in Figure 3.11 then we see the structure in the transition graph of  $\mathcal{M}_n$  that is responsible for counting the difference between the numbers of  $a$ 's and  $b$ 's that have occurred so far. We start our observations with a technical lemma which says that we find patterns from  $\mathbb{P}_{n-1}^C$  in this structure. In the following we will use  $q \xleftrightarrow[a]{b} r$  as an abbreviation for  $q \xrightarrow{a} r$  and  $r \xrightarrow{b} q$ .

**Lemma 3.42.** *For each  $n \geq 2$  there exist patterns  $p, p' \in \mathbb{P}_{n-1}^C$  such that for every DFA  $\mathcal{M} = (A, S, \delta, s_0, S')$  and all states  $r_0, \dots, r_n \in S$  with  $r_0 \xleftrightarrow[a]{b} r_1 \xleftrightarrow[a]{b} \dots \xleftrightarrow[a]{b} r_n$  it holds that  $r_0 \xrightarrow{p} r_1$  and  $r_1 \xrightarrow{p'} r_0$ .*

*Proof.* We show the lemma by induction on  $n$ . For the induction base let  $n = 2$  and define  $p =_{\text{def}} (a, p_0)$  and  $p' =_{\text{def}} (ab, p'_0, b, p'_1)$  with  $p_0 =_{\text{def}} (\varepsilon, b)$ ,  $p'_0 =_{\text{def}} (\varepsilon, a)$  and  $p'_1 =_{\text{def}} (\varepsilon, \varepsilon)$ . Obviously,  $p_0$ ,  $p'_0$  and  $p'_1$  are elements of  $\mathbb{P}_0^{\mathcal{L}}$ , and it follows that  $p, p' \in \mathbb{P}_1^{\mathcal{L}}$ .

Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA and let  $r_0, r_1, r_2$  be states of  $\mathcal{M}$  such that

$$r_0 \xleftrightarrow[b]{a} r_1 \xleftrightarrow[b]{a} r_2.$$

Then it holds that

$$r_0 \xrightarrow{a} r_1 \xrightarrow[p_0]{\text{ooo}} r_0, \quad r_1 \xrightarrow{a} r_2 \xrightarrow[p_0]{\text{ooo}} r_1 \quad \text{and} \quad r_0 \xrightarrow{a} r_1.$$

Since  $p_0$  appears at  $r_1$  this shows that  $r_0 \xrightarrow[p]{\text{ooo}} r_1$ . Furthermore, we observe

$$r_1 \xrightarrow{ab} r_1 \xrightarrow[p'_0]{\text{ooo}} r_2 \xrightarrow{b} r_1 \xrightarrow[p'_1]{\text{ooo}} r_1, \quad r_0 \xrightarrow{ab} r_0 \xrightarrow[p'_0]{\text{ooo}} r_1 \xrightarrow{b} r_0 \xrightarrow[p'_1]{\text{ooo}} r_0 \quad \text{and} \quad r_1 \xrightarrow{ab} r_1 \xrightarrow{b} r_0.$$

Together with the facts that  $p'_0$  appears at  $r_1$  and  $p'_1$  appears at  $r_0$  it follows that  $r_1 \xrightarrow[p']{\text{ooo}} r_0$ . This shows the induction base.

Now we assume that there exist  $\hat{p}, \hat{p}' \in \mathbb{P}_{n-1}^{\mathcal{L}}$  for some  $n \geq 2$  having the properties stated in the lemma and we want to show the lemma for  $n + 1$ . Define  $p =_{\text{def}} (a, \hat{p}')$  and  $p' =_{\text{def}} (ab, \hat{p}, b, \lambda(\hat{p}))$ . Since  $\lambda(\hat{p}) \in \mathbb{P}_{n-1}^{\mathcal{L}}$  by Lemma 3.9 we have  $p, p' \in \mathbb{P}_n^{\mathcal{L}}$ .

Suppose we are given a DFA  $\mathcal{M} = (A, S, \delta, s_0, S')$  and states  $r_0, r_1, \dots, r_{n+1} \in S$  with

$$r_0 \xleftrightarrow[b]{a} r_1 \xleftrightarrow[b]{a} \dots \xleftrightarrow[b]{a} r_{n+1}.$$

If we apply the induction hypothesis to the states  $r_0, \dots, r_n$  and also to the states  $r_1, \dots, r_{n+1}$  we obtain

$$r_0 \xrightarrow[\hat{p}]{\text{ooo}} r_1, \quad r_1 \xrightarrow[\hat{p}']{\text{ooo}} r_0, \quad r_1 \xrightarrow[\hat{p}]{\text{ooo}} r_2 \quad \text{and} \quad r_2 \xrightarrow[\hat{p}']{\text{ooo}} r_1. \quad (3.5)$$

It follows that  $\hat{p}$  appears at  $r_0$  and at  $r_1$ . From Lemma 3.9 we obtain

$$r_0 \xrightarrow[\lambda(\hat{p})]{\text{ooo}} r_0 \quad \text{and} \quad r_1 \xrightarrow[\lambda(\hat{p})]{\text{ooo}} r_1. \quad (3.6)$$

Let us verify that  $r_0 \xrightarrow[p]{\text{ooo}} r_1$ . Since  $r_0 \xrightarrow{a} r_1$  and  $r_1 \xrightarrow[\hat{p}']{\text{ooo}} r_0$  by (3.5), we get that  $p$  appears at  $r_0$ . Similarly, we get that  $p$  appears at  $r_1$ . Moreover,  $r_0 \xrightarrow{a} r_1$  and  $\hat{p}'$  appears at  $r_1$  by (3.5). It follows that  $r_0 \xrightarrow[p]{\text{ooo}} r_1$ .

Now we want so see that  $r_1 \xrightarrow[p']{\text{ooo}} r_0$ . From (3.5) and (3.6) we get that  $p'$  appears at  $r_1$  because

$$r_1 \xrightarrow{ab} r_1 \xrightarrow[\hat{p}]{\text{ooo}} r_2 \xrightarrow{b} r_1 \xrightarrow[\lambda(\hat{p})]{\text{ooo}} r_1.$$

Similarly, we get that  $p'$  appears at  $r_0$ . Moreover, we have  $r_1 \xrightarrow{ab} r_1 \xrightarrow{b} r_0$ . Finally we obtain  $r_1 \xrightarrow[p']{\text{ooo}} r_0$ , since  $\hat{p}$  appears at  $r_1$  by (3.5) and  $\lambda(\hat{p})$  appears at  $r_0$  by (3.6).  $\square$

From Lemma 3.42 we know that patterns from  $\mathbb{P}_{n-1}^{\mathcal{L}}$  appear at certain states in Figure 3.11. Now we exploit this to find states in  $\mathcal{M}_n$  that are connected via a pattern from  $\mathbb{P}_n^{\mathcal{L}}$ . This allows to show that  $\mathcal{M}_n$  has a pattern from  $\mathbb{P}_n^{\mathcal{L}}$ , i.e.,  $L_n \notin \mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$ .



**Lemma 3.43.** *For  $n \geq 1$  it holds that  $L_n \notin \mathcal{FP}(\mathbb{P}_n^\mathcal{L})$ .*

*Proof.* Let  $n = 1$ . Then we have to show that  $\mathcal{M}_1$  has a pattern from  $\mathbb{P}_1^\mathcal{L}$ . For this we define  $p =_{\text{def}} (ab, p_0, b, p_1)$  with  $p_0 =_{\text{def}} (\varepsilon, a)$  and  $p_1 =_{\text{def}} (\varepsilon, \varepsilon)$ . Note that  $p \in \mathbb{P}_1^\mathcal{L}$ . Now we observe that

$$r_0 \xrightarrow{ab} r_0 \xrightarrow{p_0} r_1 \xrightarrow{b} r_0 \xrightarrow{p_1} r_0, \quad r^- \xrightarrow{ab} r^- \xrightarrow{p_0} r^- \xrightarrow{b} r^- \xrightarrow{p_1} r^- \quad \text{and} \quad r_0 \xrightarrow{ab} r_0 \xrightarrow{b} r^-.$$

Together with the facts that  $p_0$  appears at  $r_0$  and  $p_1$  appears at  $r^-$  it follows that  $r_0 \xrightarrow{p} r^-$ . Since  $r_0 \xrightarrow{a} +$  and  $r^- \xrightarrow{a} -$  we obtain that  $\mathcal{M}_1$  has a pattern from  $\mathbb{P}_1^\mathcal{L}$ .

Now let  $n \geq 2$ . Then we have to show that  $\mathcal{M}_n$  has a pattern from  $\mathbb{P}_n^\mathcal{L}$ . By Lemma 3.42 there exists a pattern  $\hat{p} \in \mathbb{P}_{n-1}^\mathcal{L}$  such that  $r_0 \xrightarrow{\hat{p}} r_1$ . Note that we have also  $r^- \xrightarrow{\hat{p}} r^-$  since it holds that

$$r^- \xleftarrow{\frac{a}{b}} r^- \xleftarrow{\frac{a}{b}} \cdots \xleftarrow{\frac{a}{b}} r^-.$$

Now define  $p =_{\text{def}} (ab, \hat{p}, b, \lambda(\hat{p}))$ . Then  $p \in \mathbb{P}_n^\mathcal{L}$  and we will show that  $r_0 \xrightarrow{p} r^-$ . Note that  $\hat{p}$  appears at  $r_0$  and at  $r^-$ . So we have by Lemma 3.9 that

$$r_0 \xrightarrow{\lambda(\hat{p})} r_0 \quad \text{and} \quad r^- \xrightarrow{\lambda(\hat{p})} r^-.$$

Hence  $p$  appears at  $r_0$  and at  $r^-$  because

$$r_0 \xrightarrow{ab} r_0 \xrightarrow{\hat{p}} r_1 \xrightarrow{b} r_0 \xrightarrow{\lambda(\hat{p})} r_0 \quad \text{and} \quad r^- \xrightarrow{ab} r^- \xrightarrow{\hat{p}} r^- \xrightarrow{b} r^- \xrightarrow{\lambda(\hat{p})} r^-.$$

Moreover, we have  $r_0 \xrightarrow{ab} r_0 \xrightarrow{b} r^-$ , and we know that  $\hat{p}$  appears at  $r_0$ , and  $\lambda(\hat{p})$  appears at  $r^-$ . This shows  $r_0 \xrightarrow{p} r^-$ . Finally, from  $r_0 \xrightarrow{a^n} +$  and  $r^- \xrightarrow{a^n} -$  it follows that  $\mathcal{M}_n$  has a pattern from  $\mathbb{P}_n^\mathcal{L}$ .  $\square$

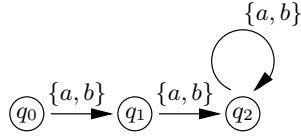
Taking together the Lemmas 3.41 and 3.43 we see that  $L_n$  separates  $\mathcal{B}_n$  and  $\mathcal{FP}(\mathbb{P}_n^\mathcal{L})$ . So we obtain the following theorem.

**Theorem 3.44.** *Let  $n \geq 1$ . Then the following holds.*

1.  $L_n \in \mathcal{B}_n \setminus \mathcal{FP}(\mathbb{P}_n^\mathcal{L})$  and  $\mathcal{B}_n \not\subseteq \mathcal{FP}(\mathbb{P}_n^\mathcal{L})$
2.  $\mathcal{FP}(\mathbb{P}_{n-1}^\mathcal{B}) \subsetneq \mathcal{FP}(\mathbb{P}_n^\mathcal{B})$  and  $\mathcal{B}_{n+1/2} \not\subseteq \mathcal{FP}(\mathbb{P}_{n-1}^\mathcal{B})$
3.  $\mathcal{FP}(\mathbb{P}_{n-1}^\mathcal{L}) \subsetneq \mathcal{FP}(\mathbb{P}_n^\mathcal{L})$  and  $\mathcal{L}_{n+1/2} \not\subseteq \mathcal{FP}(\mathbb{P}_{n-1}^\mathcal{L})$

*Proof.* From the Lemmas 3.41 and 3.43 we obtain statement 1. By Theorem 3.39.2 it follows that  $L_n \in \mathcal{FP}(\mathbb{P}_n^\mathcal{B}) \setminus \mathcal{FP}(\mathbb{P}_n^\mathcal{L})$ . By Theorem 3.38.2 we have  $\mathcal{FP}(\mathbb{P}_{n-1}^\mathcal{B}) \subseteq \mathcal{FP}(\mathbb{P}_n^\mathcal{L})$ . This implies  $L_n \in \mathcal{FP}(\mathbb{P}_n^\mathcal{B}) \setminus \mathcal{FP}(\mathbb{P}_{n-1}^\mathcal{B})$  which in turn shows  $\mathcal{FP}(\mathbb{P}_{n-1}^\mathcal{B}) \subsetneq \mathcal{FP}(\mathbb{P}_n^\mathcal{B})$  by Theorem 3.38.4. Now from  $L_n \in \mathcal{B}_n \subseteq \mathcal{B}_{n+1/2}$  we get  $L_n \in \mathcal{B}_{n+1/2} \setminus \mathcal{FP}(\mathbb{P}_{n-1}^\mathcal{B})$  and  $\mathcal{B}_{n+1/2} \not\subseteq \mathcal{FP}(\mathbb{P}_{n-1}^\mathcal{B})$ . This shows statement 2.

For the proof of statement 3 we have to distinguish between  $n = 1$  and  $n \geq 2$ . First of all we consider the case  $n \geq 2$ . From statement 1 of this theorem we know that  $L_{n-1} \in \mathcal{B}_{n-1} \setminus \mathcal{FP}(\mathbb{P}_{n-1}^\mathcal{L})$  (here we need the assumption  $n \geq 2$ ). Since  $\mathcal{B}_{n-1} \subseteq \mathcal{L}_n \subseteq \mathcal{L}_{n+1/2} \subseteq \mathcal{FP}(\mathbb{P}_n^\mathcal{L})$  (by Theorem 3.39.1) we obtain  $L_{n-1} \in \mathcal{L}_{n+1/2} \setminus \mathcal{FP}(\mathbb{P}_{n-1}^\mathcal{L})$  and  $L_{n-1} \in \mathcal{FP}(\mathbb{P}_n^\mathcal{L}) \setminus \mathcal{FP}(\mathbb{P}_{n-1}^\mathcal{L})$ . This shows statement 3 for  $n \geq 2$ .



**Fig. 3.12.** DFA  $\mathcal{M}$  with initial state  $q_0$  and accepting state  $q_1$ .  
It holds that  $L(\mathcal{M}) \in \mathcal{B}_{1/2} \setminus \mathcal{FP}(\mathbb{P}_0^{\mathcal{L}})$

Finally we have to prove statement 3 for  $n = 1$ . By Theorem 3.39.1 it suffices to construct a language  $L \in \mathcal{L}_{3/2} \setminus \mathcal{FP}(\mathbb{P}_0^{\mathcal{L}})$ . For this we consider the DFA  $\mathcal{M}$  given in Figure 3.12. The language accepted by  $\mathcal{M}$  is  $L(\mathcal{M}) = \{a, b\}$  which is in  $\mathcal{B}_{1/2} \subseteq \mathcal{L}_{3/2}$ . Moreover, for  $p =_{\text{def}} (\varepsilon, a)$  we have

$$q_0 \xrightarrow{a} q_1, \quad q_1 \xrightarrow{p} q_2, \quad q_1 \xrightarrow{\varepsilon} + \quad \text{and} \quad q_2 \xrightarrow{\varepsilon} -.$$

Therefore,  $\mathcal{M}$  has a pattern from  $\mathbb{P}_0^{\mathcal{L}}$  which shows that  $L \notin \mathcal{FP}(\mathbb{P}_0^{\mathcal{L}})$ . This proves statement 3 for the case  $n = 1$ .  $\square$

**Remark 3.45.** Suppose we deal with some alphabet  $A$  such that  $|A| > 2$ , e.g.,  $A = \{a, b, c_1, \dots, c_n\}$  for some  $n \geq 1$ . If we define  $\mathcal{M}_n$  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 we can show Lemma 3.43. This means on the language side that we intersect the expressions for  $L_n$  with  $\{a, b\}^+ = A^+ \setminus \bigcup_{1 \leq i \leq n} A^* c_i A^* \in \text{co}\mathcal{B}_{1/2}$ . The latter does not increase the dot-depth since  $L_n \in \mathcal{B}_n$  which is a Boolean algebra that contains  $\text{co}\mathcal{B}_{1/2}$ . Together this allows to prove Theorem 3.44 also in case of a larger alphabet.

### 3.7.4 The Classes $\mathcal{FP}(\mathbb{P}_n^{\mathcal{B}})$ and $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$ are Decidable

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

**Theorem 3.46.** *Fix some  $n \geq 0$ . On input of a DFA  $\mathcal{M}$  it is decidable in nondeterministic logarithmic space whether  $L(\mathcal{M})$  is in  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{B}})$  (respectively,  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$ ).*

*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 classes of initial patterns the problems  $\text{PATTERN}_{0,k}^{\mathcal{L}}$  and  $\text{PATTERN}_{0,k}^{\mathcal{B}}$  are just reachability problems very similar to  $\text{REACH}_k$ , which can be solved in NL (see Lemma 3.32). Now the theorem follows from Theorem 3.35.  $\square$

Since membership to SF is decidable, this yields an algorithm to determine the minimal  $n$  such that  $L(\mathcal{M})$  is in  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{B}})$  (respectively,  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$ ). Moreover, note that although the single classes  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$  and  $\mathcal{FP}(\mathbb{P}_n^{\mathcal{B}})$  are decidable in NL, the decision problem for SF is known to be PSPACE-complete [CH91].

### 3.7.5 Lower Bounds for the Dot-Depth Problem

We summarize the inclusion structure of concatenation hierarchies and forbidden-pattern hierarchies in Figure 3.13 where inclusions hold from bottom to top. Observe the structural similarities of these hierarchies. In fact we will specify the picture in chapter 4 (cf. Figure 4.7), i.e., we will show that the levels 0 and 1 of the forbidden-pattern hierarchies coincide with the levels 1/2 and 3/2 of the respective concatenation hierarchies. Moreover, one can show that even  $\mathcal{L}_{5/2} = \mathcal{FP}(\mathbb{P}_2^c)$  if we consider a two-letter alphabet [GS00b].

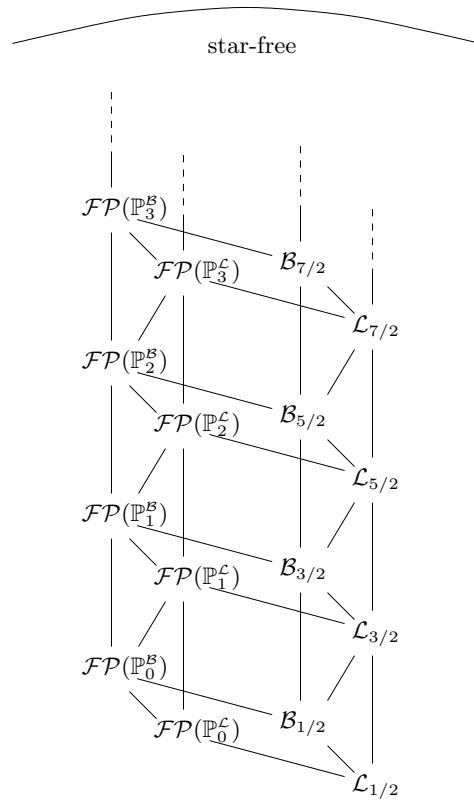


Fig. 3.13. Concatenation hierarchies and forbidden-pattern hierarchies

In fact, the inclusion  $\mathcal{B}_{n+1/2} \subseteq \mathcal{FP}(\mathbb{P}_n^B)$  establishes a lower bound algorithm for the dot-depth of a given language. This follows from the fact that we can determine the minimal  $n$  such that a given language is in  $\mathcal{FP}(\mathbb{P}_n^B)$ . Moreover, if we look at the Theorems 3.44.2 and 3.39.2 we see that the pattern class  $\mathcal{FP}(\mathbb{P}_n^B)$  captures  $\mathcal{B}_{n+1/2}$  but not  $\mathcal{B}_{n+3/2}$ . This indicates that the forbidden-pattern classes are not ‘too big’.

### 3.8 Summary and Discussion

We started this chapter with some observations of regularities in known forbidden-pattern characterizations for concatenation hierarchies. From this we derived an iteration rule IT

which led to hierarchies of pattern classes. Then we considered the classes of languages being accepted by DFAs that does not have the patterns from these classes.

In Theorem 3.16 we compared the polynomial closure operation (which is used in the definition of concatenation hierarchies) with our iteration rule IT (which provides decidable classes of languages). Note that our result  $\text{Pol}(\text{co}\mathcal{FP}(\mathbb{P}_n^x)) \subseteq \mathcal{FP}(\mathbb{P}_{n+1}^x)$  generalizes the usually easier to prove inclusion in forbidden-pattern characterizations, and of course we are interested in the reverse inclusion.

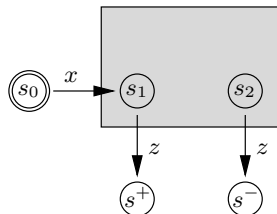
So far we were able to prove that the classes  $\mathcal{FP}(\mathbb{P}_n^c)$  and  $\mathcal{FP}(\mathbb{P}_n^s)$  form strict and decidable hierarchies exhausting the class of starfree languages. Moreover, these classes contain the respective levels of the STH and DDH, the emerging hierarchies show the same inclusion structure as it is known from concatenation hierarchies, and we gave some evidence that our pattern classes are not ‘too big’.

In the next chapter we show that for some lower levels the forbidden-pattern classes coincide with the classes of concatenation hierarchies. In particular we obtain a forbidden-pattern characterization for level 3/2 of the DDH.

## 4. Decidability Results for the DDH and STH

The classes of the DDH and STH formalize the famous dot-depth problem in terms of the decidability of their membership problems. In the last 30 years it turned out that these decidability questions are extremely difficult. So up to now, only some lower levels of both hierarchies are known to be decidable. In this chapter we restrict ourselves to levels  $n + 1/2$  and we (re)prove decidability results for these levels. More precisely, we prove the decidability—in terms of forbidden-pattern characterizations—of the levels  $1/2$  and  $3/2$  of both hierarchies. Furthermore, we also show the decidability of the Boolean hierarchies over the levels  $1/2$  of the STH and DDH.

To obtain these results we use a technique which is based on *word extensions*. Their definitions depend on the considered level (in our case  $1/2$  or  $3/2$ ), on a parameter  $k \geq 0$  and on a DFA  $\mathcal{M}$ . With this technique it is possible to treat the classes  $\mathcal{L}_{1/2}$ ,  $\mathcal{B}_{1/2}$ ,  $\mathcal{L}_{3/2}$  and  $\mathcal{B}_{3/2}$  in a very similar way. Moreover, the word extensions for the levels  $1/2$  can be used to prove the decidability of the corresponding Boolean hierarchies.



**Fig. 4.1.** General structure of forbidden-patterns.

In Figure 4.1 we recall the general structure of forbidden-patterns from chapter 3. In a sense, our word extensions are based on such forbidden-patterns. By their construction it is ensured that a word  $w$  can only be extended in the following special way: If we consider the path in a DFA induced by  $w$  then we may extend this word only at positions where we have reached a state similar to  $s_1$ , and we may insert only such words that lead to a state similar to  $s_2$ . We will exploit this connection between patterns and word extensions in the following proofs.

At this point we want to make a remark concerning our notations:  $\preceq_{\mathcal{M}}^{0,k}$  will denote the extensions that correspond to level  $1/2$ , and  $\preceq_{\mathcal{M}}^{1,k}$  will denote the extensions that correspond to level  $3/2$ . In this notation we use 0 instead of  $1/2$  and 1 instead of  $3/2$  because we want to emphasize on the connection between the word extensions and the forbidden-pattern classes. Later we will see that  $\mathcal{B}_{1/2} = \mathcal{FP}(\mathbb{P}_0^{\mathcal{B}})$  and  $\mathcal{B}_{3/2} = \mathcal{FP}(\mathbb{P}_1^{\mathcal{B}})$ . This means that  $\preceq_{\mathcal{M}}^{0,k}$

corresponds to  $\mathcal{FP}(\mathbb{P}_0^{\mathcal{B}})$  and  $\preceq_{\mathcal{M}}^{1,k}$  corresponds to  $\mathcal{FP}(\mathbb{P}_1^{\mathcal{B}})$ . Throughout this chapter we treat the DDH and the STH in parallel, i.e., all theorems will be proved for both hierarchies.

## 4.1 The Levels 1/2

Here we consider the levels 1/2 of the DDH and STH, and we prove effective forbidden-pattern characterizations for them. These results were first shown in [Arf91, PW97]. However, for reasons of methodology we give reproofs that use the technique of word extensions. In section 4.3 we apply the same technique to the levels 3/2, and we obtain effective forbidden-pattern characterizations for these levels.

We start with the definition of the word extensions  $\preceq_{\mathcal{M}}^{0,k}$  which can be also considered as binary, reflexive, transitive and antisymmetric relations on the set of words. Then in subsection 4.1.2 we show that the  $\preceq_{\mathcal{M}}^{0,k}$  upward closure of a nonempty word (in other words, the  $\preceq_{\mathcal{M}}^{0,k}$  co-ideal generated by a nonempty word) is in  $\mathcal{B}_{1/2}$ . In subsection 4.1.3 we prove that the set of words together with  $\preceq_{\mathcal{M}}^{0,k}$  is a well partial ordered set. In particular, all  $\preceq_{\mathcal{M}}^{0,k}$  co-ideals are finitely generated. Together with the result of subsection 4.1.2 this implies that  $\preceq_{\mathcal{M}}^{0,k}$  co-ideals are in  $\mathcal{B}_{1/2}$ . Finally, we show that languages from the forbidden-pattern class  $\mathcal{FP}(\mathbb{P}_0^{\mathcal{B}})$  are  $\preceq_{\mathcal{M}}^{0,k}$  co-ideals, and therefore we obtain  $\mathcal{FP}(\mathbb{P}_0^{\mathcal{B}}) \subseteq \mathcal{B}_{1/2}$ . Since the reverse inclusion is known from the pattern theory in chapter 3, we obtain the effective forbidden-pattern characterization  $\mathcal{FP}(\mathbb{P}_0^{\mathcal{B}}) = \mathcal{B}_{1/2}$  which shows that the membership problem for  $\mathcal{B}_{1/2}$  is decidable in nondeterministic logarithmic space. We show analogous results for the STH.

### 4.1.1 Definition of $\preceq_{\mathcal{M}}^{0,k}$ Word Extensions

In section 1.4 we introduced  $\prec_v^{0,k}$  word extensions. By their definition they are such that some factor is inserted at a certain position in the initial word. Here ‘‘certain position’’ means a position where the word  $v$  appears. Therefore, this word  $v$  is also called a context word. In this subsection we introduce  $\preceq_{\mathcal{M}}^{0,k}$  extensions. They are defined in such a way that  $x \preceq_{\mathcal{M}}^{0,k} y$  means that the word  $y$  results from the word  $x$  by a sequence of extensions  $\prec_{v_1}^{0,k}, \prec_{v_2}^{0,k}, \dots, \prec_{v_m}^{0,k}$ . Here it is important that we do not fix the context word but we allow several words  $v_0, v_1, \dots, v_m$ . However, only context words of a certain length are allowed. Therefore, we define the set of possible context words.

**Definition 4.1.** For a DFA  $\mathcal{M}$  and  $k \geq 0$  let  $W_{\mathcal{M}}^{0,k} =_{\text{def}} A^k$  be the set of context words.

Actually the definition of  $W_{\mathcal{M}}^{0,k}$  does not depend on the DFA  $\mathcal{M}$ , i.e.,  $W_{\mathcal{M}_1}^{0,k} = W_{\mathcal{M}_2}^{0,k}$  for all DFAs  $\mathcal{M}_1, \mathcal{M}_2$ . However, we use the notation above in order to stress the similarities in the approaches for the levels 1/2 and 3/2 (in section 4.3 we will see that for level 3/2 the set of context words depends on the DFA). Now we define  $\preceq_{\mathcal{M}}^{0,k}$  extensions as sequences of  $\prec_v^{0,k}$  extensions with  $v \in W_{\mathcal{M}}^{0,k}$ .

**Definition 4.2.** Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $y, y' \in A^*$ .

$$y \preceq_{\mathcal{M}}^{0,k} y' \iff_{\text{def}} \text{there exist an } m \geq 0, \text{ words } x_0, \dots, x_m \in A^* \text{ and } v_1, \dots, v_m \in W_{\mathcal{M}}^{0,k} \\ \text{such that } y = x_0 \prec_{v_1}^{0,k} x_1 \prec_{v_2}^{0,k} \dots \prec_{v_m}^{0,k} x_m = y'$$

From this definition it follows that  $\preceq_{\mathcal{M}}^{0,k}$  is reflexive, transitive and antisymmetric. We have already seen that the set of context words  $W_{\mathcal{M}}^{0,k}$  does not depend on the DFA  $\mathcal{M}$ . This carries over to  $\preceq_{\mathcal{M}}^{0,k}$  extensions.

**Proposition 4.3.** *For  $k \geq 0$ , DFAs  $\mathcal{M}_1, \mathcal{M}_2$  and  $y, y' \in A^*$  it holds that*

$$y \preceq_{\mathcal{M}_1}^{0,k} y' \iff y \preceq_{\mathcal{M}_2}^{0,k} y'.$$

*Proof.* This follows from Definition 4.2 and the fact that  $W_{\mathcal{M}_1}^{0,k} = W_{\mathcal{M}_2}^{0,k} = A^k$ .  $\square$

Next we prove some basic facts about  $\preceq_{\mathcal{M}}^{0,k}$ . In particular we show that  $\preceq_{\mathcal{M}}^{0,k}$  are stable word extensions which preserve the  $k$ -prefix and the  $k$ -suffix.

**Proposition 4.4.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$ ,  $v \in W_{\mathcal{M}}^{0,k}$  and  $w, w' \in A^*$ . Then the following holds.*

1. *If  $w \preceq_{\mathcal{M}}^{0,k} w'$  then  $\mathbf{p}_k(w) = \mathbf{p}_k(w')$  and  $\mathbf{s}_k(w) = \mathbf{s}_k(w')$ .*
2. *If  $w \preceq_{\mathcal{M}}^{0,k} w'$  then  $xwz \preceq_{\mathcal{M}}^{0,k} xw'z$  for all  $x, z \in A^*$ .*
3. *If  $v \preceq_{\mathcal{M}}^{0,k} w$  then  $v = w$  or  $v \prec_v^{0,k} w$ .*

*Proof.* Trivially, if  $w = w'$  then the statements 1 and 2 hold. If  $w \neq w'$  then by definition there exist an  $m \geq 1$ , words  $x_0, \dots, x_m \in A^*$  and  $v_1, \dots, v_m \in W_{\mathcal{M}}^{0,k}$  such that  $w = x_0 \prec_{v_1}^{0,k} x_1 \prec_{v_2}^{0,k} \dots \prec_{v_m}^{0,k} x_m = w'$ . The statements 1 and 2 follow from Proposition 1.19 since  $|v_i| = k$  for  $1 \leq i \leq m$ .

For statement 3 assume that  $v \preceq_{\mathcal{M}}^{0,k} w$ . If  $v = w$  then we are done. Otherwise, statement 1 and Definition 4.2 imply that  $w \in vA^{\geq k+1}v$ . This shows  $v \prec_v^{0,k} w$ .  $\square$

**Proposition 4.5.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $y_1, y_2, z \in A^*$  with  $y_1y_2 \preceq_{\mathcal{M}}^{0,k} z$ . Then there exist words  $z_1, z_2 \in A^*$  such that  $z = z_1z_2$ ,  $|y_1| \leq |z_1|$  and  $y_2 \preceq_{\mathcal{M}}^{0,k} z_2$ .*

*Proof.* By Definition 4.2 it suffices to show the proposition for  $\prec_v^{0,k}$ . The following claim achieves this.

**Claim.** *Let  $y_1, y_2, z \in A^*$  and  $v \in A^k$  with  $y_1y_2 \prec_v^{0,k} z$ . Then there exist words  $z_1, z_2 \in A^*$  with  $z = z_1z_2$  and  $|y_1| \leq |z_1|$  such that either  $y_2 = z_2$  or  $y_2 \prec_v^{0,k} z_2$ .*

For  $y =_{\text{def}} y_1y_2$  there exist words  $y'_1, y'_2 \in A^*$  and  $u \in A^{\geq k+1}$  such that  $y = y'_1vy'_2$  and  $z = y'_1vuvy'_2$ . Now we compare the decompositions  $y = y_1y_2$  and  $y = y'_1vy'_2$ .

If  $y_2$  is a suffix of  $vy'_2$  (i.e.,  $|y_2| \leq |vy'_2|$ ) then there exists some  $w \in A^*$  with  $vy'_2 = wy_2$ . It follows that  $y_1 = y'_1w$ . Therefore, with  $z_1 =_{\text{def}} y'_1vuw$  and  $z_2 =_{\text{def}} y_2$  we obtain  $z = z_1z_2$ ,  $|z_1| \geq |y'_1w| = |y_1|$  and  $y_2 = z_2$ .

If  $y_2$  is not a suffix of  $vy'_2$  (i.e.,  $|y_2| > |vy'_2|$ ) then  $y_2 = wvy'_2$  for some  $w \in A^+$ . It follows that  $y'_1 = y_1w$ . With  $z_1 =_{\text{def}} y_1$  and  $z_2 =_{\text{def}} wvvy'_2$  we get  $z = z_1z_2$  and  $|z_1| = |y_1|$ . Moreover, it holds that  $y_2 = wvy'_2 \prec_v^{0,k} wvvy'_2 = z_2$ .  $\square$

### 4.1.2 The $\preceq_{\mathcal{M}}^{0,k}$ Upward Closure of a Word

For a nonempty word  $y$  we want to show that the  $\preceq_{\mathcal{M}}^{0,0}$  upward closure of  $y$  is in  $\mathcal{L}_{1/2}$ , and that the  $\preceq_{\mathcal{M}}^{0,k}$  upward closure of  $y$  is in  $\mathcal{B}_{1/2}$  for  $k \geq 0$ . The idea is similar to that in subsection 1.4.3 where we considered the  $\prec_v^{1,k}$  upward closure of nonempty words: If  $y \preceq_{\mathcal{M}}^{0,k} y'$  then by definition  $y'$  emerges from  $y$  by a sequence of  $\prec_v^{0,k}$  extensions. Remember that a single  $\prec_v^{0,k}$  extension is such that a given word is modified by inserting some letters at exactly one position in this word. The following lemma shows that in the sequence leading from  $y$  to  $y'$  one can trace back these positions. This yields a list of positions in  $y$  that can be used to transform  $y$  into  $y'$  in a single step (where  $\prec_v^{0,k}$  extensions are carried out, in parallel, at several positions in  $y$ ). Since the number of these positions is  $\leq |y| + 1$  there exists a sequence of length  $\leq |y| + 1$  leading from  $y$  to  $y'$ . At the end this will show  $\langle y \rangle_{\preceq_{\mathcal{M}}^{0,0}} \in \mathcal{L}_{1/2}$  and  $\langle y \rangle_{\preceq_{\mathcal{M}}^{0,k}} \in \mathcal{B}_{1/2}$ .

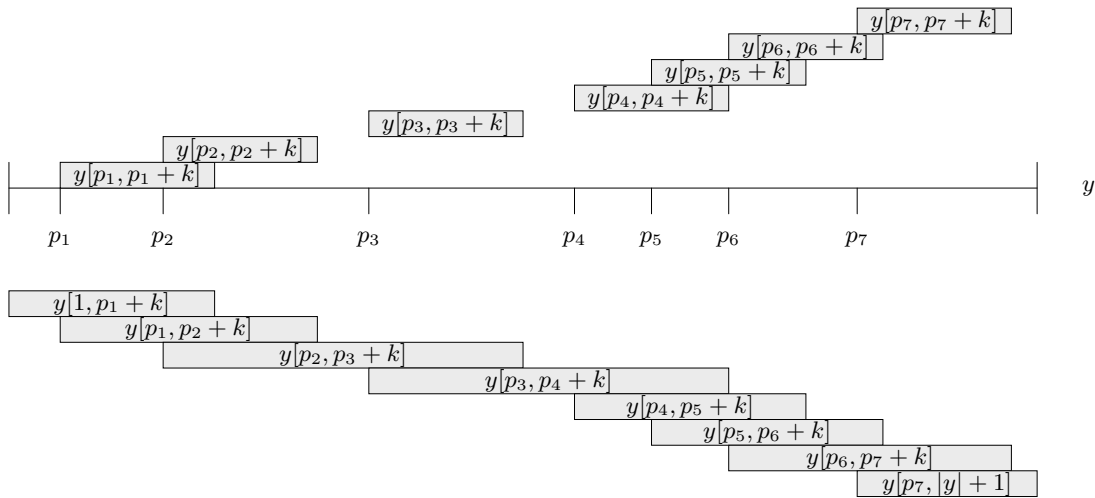
Note that the following lemma also shows a possibility to define  $\preceq_{\mathcal{M}}^{0,k}$  extensions directly, i.e., without using  $\prec_v^{0,k}$  chains. The proof below is an adapted version of the proof of Lemma 1.26 where we showed a similar result for  $\prec_v^{1,k}$  extensions.

**Lemma 4.6.** *For every DFA  $\mathcal{M}$ ,  $k \geq 0$  and  $y \in A^*$  it holds that*

$$\langle y \rangle_{\preceq_{\mathcal{M}}^{0,k}} = \{y\} \cup \bigcup y[1, p_1 + k] \cdot A^{*\geq k+1} \cdot y[p_1, p_2 + k] \cdot A^{*\geq k+1} \cdot y[p_2, p_3 + k] \cdot \dots \cdot A^{*\geq k+1} \cdot y[p_{m-1}, p_m + k] \cdot A^{*\geq k+1} \cdot y[p_m, |y| + 1] \quad (4.1)$$

where the union ranges over all  $m \geq 1$  and all positions  $1 \leq p_1 < \dots < p_m \leq |y| - k + 1$ .

*Proof.* The following picture illustrates the idea of the union in the lemma. It shows the factors that emerge when we consider the positions  $p_1, p_2, \dots, p_7$  in  $y$ . The upper part of the picture shows the blocks of length  $k$  that have to be doubled when making  $\prec_v^{0,k}$  extensions at the positions  $p_i$ . In the lower part we see the factors of  $y$  that remain connected. Note that in the lower part, neighboring factors overlap in exactly  $k$  letters.





In the proof we denote the right-hand side of (4.1) by  $L$ . At first we show  $\langle y \rangle_{\underset{\mathcal{M}}{\geq} 0, k} \subseteq L$ . For this we assume that  $\langle y \rangle_{\underset{\mathcal{M}}{\geq} 0, k} \not\subseteq L$ , this will lead to a contradiction. Since at least  $y$  is in  $L$ , there exist words  $w, w' \in A^*$  and  $v \in W_{\mathcal{M}}^{0, k}$  such that  $w \in L$ ,  $w' \notin L$  and  $w \underset{v}{<}^{0, k} w'$ . Hence there exist words  $x, z \in A^*$  and  $u \in A^{*\geq k+1}$  such that  $w = xvz$  and  $w' = xvuvz$ .

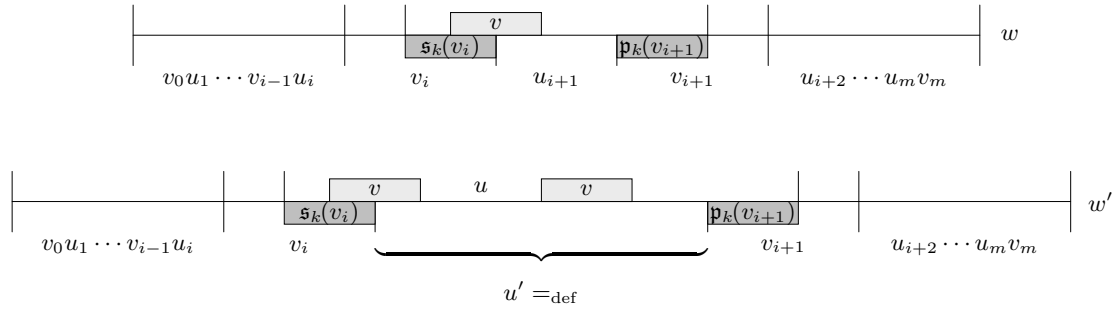
If  $w = y$  then with  $p_1 =_{\text{def}} |x| + 1$  we obtain  $y[1, p_1 + k] = xv$  and  $y[p_1, |y| + 1] = vz$ . Therefore, we get  $w' \in y[1, p_1 + k] \cdot A^{*\geq k+1} \cdot y[p_1, |y| + 1] \subseteq L$  which contradicts our assumption.

Assume now  $w \neq y$ . Then there exist an  $m \geq 1$ , positions  $1 \leq p_1 < \dots < p_m \leq |y| - k + 1$  and words  $u_1, \dots, u_m \in A^{*\geq k+1}$  such that

$$w = \underbrace{y[1, p_1 + k]}_{v_0 = \text{def}} \cdot u_1 \cdot \underbrace{y[p_1, p_2 + k]}_{v_1 = \text{def}} \cdot u_2 \cdot \underbrace{y[p_2, p_3 + k]}_{v_2 = \text{def}} \cdot \dots \cdot u_m \cdot \underbrace{y[p_m, |y| + 1]}_{v_m = \text{def}}. \quad (4.2)$$

For  $0 \leq i \leq m$  define  $v_i$  as above and note that  $|v_i| \geq k$  (for  $1 \leq i \leq m - 1$  it even holds that  $|v_i| \geq k + 1$ ). Now we compare the decompositions (4.2) and  $w = xvz$ .

**Case 1:** Assume that the factor  $v$  of the decomposition  $w = xvz$  is contained in some factor  $\mathfrak{s}_k(v_i) \cdot u_{i+1} \cdot \mathfrak{p}_k(v_{i+1})$  of the decomposition (4.2). Then we have the following situation.



Define  $u'$  as in the picture above and observe that  $w' = v_0 u_1 v_1 \dots u_i v_i u' v_{i+1} \dots u_m v_m$ . Since  $|u'| \geq k + 1$  we get

$$w' \in y[1, p_1 + k] \cdot A^{*\geq k+1} \cdot y[p_1, p_2 + k] \cdot A^{*\geq k+1} \cdot y[p_2, p_3 + k] \cdot \dots \cdot A^{*\geq k+1} \cdot y[p_m, |y| + 1].$$

This contradicts our assumption  $w' \notin L$ .

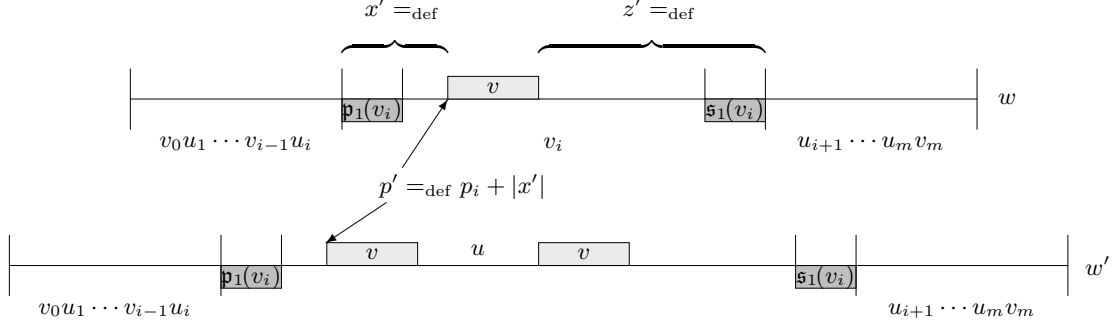
**Case 2:** Assume now that the factor  $v$  of the decomposition  $w = xvz$  is *not* contained in some factor  $\mathfrak{s}_k(v_i) \cdot u_{i+1} \cdot \mathfrak{p}_k(v_{i+1})$  of the decomposition (4.2). Since all  $\mathfrak{s}_k(v_i) \cdot u_{i+1} \cdot \mathfrak{p}_k(v_{i+1})$  are longer than  $k = |v|$ , it must be that one of the following subcases occurs.

*Case 2a:*  $v$  is contained in some factor  $A^{-1}v_i A^{-1}$  for  $1 \leq i \leq m - 1$

*Case 2b:*  $v$  is contained in  $v_0 A^{-1}$

*Case 2c:*  $v$  is contained in  $A^{-1}v_m$

We will only treat Case 2a, the other cases are analogous. Hence our current situation is as follows.



Define  $p'$  as in the picture. Since  $x', z'$  are nonempty we have  $1 \leq |x'| \leq |v_i| - k - 1$ . Together with  $|v_i| = p_{i+1} - p_i + k$  this implies  $1 \leq |x'| \leq p_{i+1} - p_i - 1$ , and therefore  $p_i < p' < p_{i+1}$ . We obtain  $y[p', p' + k] = v$  because  $v_i = y[p_i, p_{i+1} + k]$  and  $v_i[|x'| + 1, |x'| + 1 + k] = v$ . Now consider the term of the union in  $L$  that takes the positions

$$1 \leq p_1 < \dots < p_i < p' < p_{i+1} < \dots < p_m \leq |y| - k + 1$$

into account. Since  $y[p_i, p' + k] = x'v$  and  $y[p', p_{i+1} + k] = vz'$  this term is equal to

$$\begin{aligned} L' =_{\text{def}} & v_0 \cdot A^{*\geq k+1} \cdot v_1 \cdot A^{*\geq k+1} \dots v_{i-1} \cdot A^{*\geq k+1} \cdot x'v \cdot A^{*\geq k+1} \cdot vz' \cdot A^{*\geq k+1} \cdot v_{i+1} \\ & \cdot A^{*\geq k+1} \cdot v_{i+2} \dots A^{*\geq k+1} \cdot v_m. \end{aligned}$$

Since  $w' = v_0 u_1 v_1 u_2 \dots v_{i-1} u_i \cdot x'v \cdot u \cdot vz' \cdot u_{i+1} v_{i+1} u_{i+2} v_{i+2} \dots u_m v_m$  we get  $w' \in L' \subseteq L$  which contradicts our assumption.

So in all considered cases we get contradictions. Therefore, our assumption was false and it follows that  $\langle y \rangle_{\stackrel{0,k}{\succeq} \mathcal{M}} \subseteq L$ . So we have shown that the left-hand side is a subset of the right-hand side in (4.1).

We turn to the proof of the reverse inclusion. Clearly, it holds that  $\langle y \rangle_{\stackrel{0,k}{\succeq} \mathcal{M}} \supseteq \{y\}$ . So let  $m \geq 1$  and choose positions  $1 \leq p_1 < \dots < p_m \leq |y| - k + 1$ . In order to show  $y' \in \langle y \rangle_{\stackrel{0,k}{\succeq} \mathcal{M}}$  for all  $y' \in L \setminus \{y\}$  we choose arbitrary  $u_1, \dots, u_m \in A^{*\geq k+1}$  and let

$$y' =_{\text{def}} y[1, p_1 + k] \cdot u_1 \cdot y[p_1, p_2 + k] \cdot u_2 \cdot y[p_2, p_3 + k] \dots u_{m-1} \cdot y[p_{m-1}, p_m + k] \cdot u_m \cdot y[p_m, |y| + 1].$$

Since  $y = y[1, p_m] \cdot y[p_m, |y| + 1]$  we get  $y <_{v_m}^{0,k} y_m$  for  $v_m =_{\text{def}} \mathfrak{p}_k(y[p_m, |y| + 1])$  and

$$y_m =_{\text{def}} y[1, p_m] \cdot v_m \cdot u_m \cdot y[p_m, |y| + 1].$$

Since  $y_m$  can also be written as  $y_m = y[1, p_{m-1}] \cdot y[p_{m-1}, p_m + k] \cdot u_m \cdot y[p_m, |y| + 1]$  we obtain  $y_m <_{v_{m-1}}^{0,k} y_{m-1}$  for  $v_{m-1} =_{\text{def}} \mathfrak{p}_k(y[p_{m-1}, p_m + k])$  and

$$y_{m-1} =_{\text{def}} y[1, p_{m-1}] \cdot v_{m-1} \cdot u_{m-1} \cdot y[p_{m-1}, p_m + k] \cdot u_m \cdot y[p_m, |y| + 1].$$

We continue this argumentation until we obtain  $y_2 \prec_{v_1}^{0,k} y_1$  for  $v_1 =_{\text{def}} \mathfrak{p}_k(y[p_1, p_2 + k])$  and  $y_1 =_{\text{def}} y[1, p_1] \cdot v_1 \cdot u_1 \cdot y[p_1, p_2 + k] \cdot u_2 \cdot y[p_2, p_3 + k] \cdots u_{m-1} \cdot y[p_{m-1}, p_m + k] \cdot u_m \cdot y[p_m, |y| + 1]$ .

Since  $y[1, p_1] \cdot v_1 = y[1, p_1 + k]$  we have  $y_1 = y'$  and  $y \prec_{v_m}^{0,k} y_m \prec_{v_{m-1}}^{0,k} \cdots \prec_{v_2}^{0,k} y_2 \prec_{v_1}^{0,k} y'$ . This shows  $y' \in \langle y \rangle_{\preceq_{\mathcal{M}}^{0,k}}$  and it follows that in (4.1) the right-hand side is a subset of the left-hand side.  $\square$

With the decomposition of Lemma 4.6 at hand we can prove that the  $\preceq_{\mathcal{M}}^{0,k}$  upward closure of a nonempty word is in  $\mathcal{B}_{1/2}$  (respectively,  $\mathcal{L}_{1/2}$  for  $k = 0$ ).

**Theorem 4.7.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA and  $k \geq 0$ . Then for all  $y \in A^+$  it holds that  $\langle y \rangle_{\preceq_{\mathcal{M}}^{0,0}} \in \mathcal{L}_{1/2}$  and  $\langle y \rangle_{\preceq_{\mathcal{M}}^{0,k}} \in \mathcal{B}_{1/2}$ .*

*Proof.* Let  $y \in A^+$ ,  $n =_{\text{def}} |y|$  and  $n' =_{\text{def}} |y| - k + 1$ . Apply Lemma 4.6 and observe that (i) the union there is finite and (ii)  $A^{*\geq k+1}$  can be written as the (finite) union of all  $wA^+$  for  $w \in A^k$ . Hence  $\langle y \rangle_{\preceq_{\mathcal{M}}^{0,k}} \setminus \{y\}$  is a finite union of languages of the form  $u_0A^+u_1 \cdots A^+u_n$  for  $n \geq 1$  and  $u_i \in A^*$ . Since  $n \geq 1$  and  $A^+ = \bigcup_{a \in A} (\{a\} \cup aA^+)$  this can be easily transformed to a finite union of languages of the form  $u_0A^+u_1 \cdots A^+u_n$  for  $n \geq 0$  and  $u_i \in A^+$ . This shows that  $\langle y \rangle_{\preceq_{\mathcal{M}}^{0,k}} \in \mathcal{B}_{1/2}$ .

Now we consider the case  $k = 0$  and set  $n =_{\text{def}} |y|$ . Lemma 4.6 says that

$$\langle y \rangle_{\preceq_{\mathcal{M}}^{0,0}} = \{y\} \cup \bigcup y[1, p_1] \cdot A^+ \cdot y[p_1, p_2] \cdot A^+ \cdot y[p_2, p_3] \cdots A^+ \cdot y[p_m, n + 1]$$

where the union ranges over all  $m \geq 1$  and all positions  $1 \leq p_1 < \cdots < p_m \leq n + 1$ . In other words, with the  $p_i$  we guess the positions in  $y$  where we insert one or more letters, and at all other positions we insert zero letters. This is equivalent to inserting words from  $A^*$  at all possible positions (i.e.,  $m =_{\text{def}} n + 1, p_1 =_{\text{def}} 1, p_2 =_{\text{def}} 2, \dots, p_m =_{\text{def}} m$ ). It follows that

$$\langle y \rangle_{\preceq_{\mathcal{M}}^{0,0}} = \underbrace{y[1, 1]}_{=\varepsilon} \cdot A^* \cdot y[1, 2] \cdot A^* \cdot y[2, 3] \cdots A^* \cdot y[n - 1, n] \cdot A^* \cdot y[n, n + 1] \cdot \underbrace{A^*}_{=\varepsilon} \cdot y[n + 1, n + 1].$$

This shows  $\langle y \rangle_{\preceq_{\mathcal{M}}^{0,0}} \in \mathcal{L}_{1/2}$ .  $\square$

### 4.1.3 $\preceq_{\mathcal{M}}^{0,k}$ Co-Ideals are Finitely Generated

In this subsection we show a fundamental property of  $\preceq_{\mathcal{M}}^{0,k}$ : The set of words together with  $\preceq_{\mathcal{M}}^{0,k}$  is a well partial ordered set. In particular this means that all  $\preceq_{\mathcal{M}}^{0,k}$  co-ideals are finitely generated. The proof we give below is based on an idea from [SS83] where the usual subword relation  $\preceq$  is considered.

We show here that in  $A^*$  there exists neither an infinite strictly descending  $\preceq_{\mathcal{M}}^{0,k}$  chain, nor an infinite set of pairwise incomparable elements with respect to  $\preceq_{\mathcal{M}}^{0,k}$ . In case  $k = 0$  we encounter the subword relation  $\preceq$  (also called division ordering) with its fundamental theorem from Higman [Hig52].

**Theorem 4.8.** *Let  $\mathcal{M}$  be a DFA and  $k \geq 0$ . It holds that  $(A^*, \preceq_{\mathcal{M}}^{0,k})$  is a wpos.*

*Proof.* For words  $x, y \in A^*$  define  $x \preceq_{\mathcal{M}}^{0,k} y$  if and only if  $y = y_1 y_2$  and  $x \preceq_{\mathcal{M}}^{0,k} y_2$  for suitable  $y_1 \in A^{\geq 2k+1} \cup \{\varepsilon\}$  and  $y_2 \in A^*$ . Observe that  $\preceq_{\mathcal{M}}^{0,k}$  is reflexive and that  $\preceq_{\mathcal{M}}^{0,k}$  is a refinement of  $\preceq_{\mathcal{M}}^{0,k}$ . Now let us see that  $\preceq_{\mathcal{M}}^{0,k}$  is even transitive. Let  $x, y, z \in A^*$  with  $x \preceq_{\mathcal{M}}^{0,k} y$  and  $y \preceq_{\mathcal{M}}^{0,k} z$ . Hence we can choose  $y_1, z_1 \in A^{\geq 2k+1} \cup \{\varepsilon\}$  and  $y_2, z_2 \in A^*$  such that  $y = y_1 y_2$ ,  $z = z_1 z_2$ ,  $x \preceq_{\mathcal{M}}^{0,k} y_2$  and  $y \preceq_{\mathcal{M}}^{0,k} z_2$ . We apply Proposition 4.5 to  $y_1 y_2 \preceq_{\mathcal{M}}^{0,k} z_2$  and obtain words  $z_{2,1}, z_{2,2} \in A^*$  such that  $z_2 = z_{2,1} z_{2,2}$ ,  $|y_1| \leq |z_{2,1}|$  and  $y_2 \preceq_{\mathcal{M}}^{0,k} z_{2,2}$ . Hence, for  $z'_1 =_{\text{def}} z_1 z_{2,1}$  and  $z'_2 =_{\text{def}} z_{2,2}$  we get  $z = z'_1 z'_2$ ,  $|y_1| \leq |z'_1|$  and  $x \preceq_{\mathcal{M}}^{0,k} y_2 \preceq_{\mathcal{M}}^{0,k} z'_2$ . If  $z'_1 \in A^{\geq 2k+1} \cup \{\varepsilon\}$  then  $x \preceq_{\mathcal{M}}^{0,k} z$  and we are done. Otherwise we have  $1 \leq |z'_1| \leq 2k$  and it follows that  $|z_1| \leq 2k$  and  $|z_{2,1}| \leq 2k$ . Since  $|y_1| \leq |z_{2,1}|$  we have also  $|y_1| \leq 2k$ . Together with  $y_1, z_1 \in A^{\geq 2k+1} \cup \{\varepsilon\}$  this implies  $y_1 = z_1 = \varepsilon$ . Finally, from  $x \preceq_{\mathcal{M}}^{0,k} y_2$  and  $y \preceq_{\mathcal{M}}^{0,k} z_2$  we obtain  $x \preceq_{\mathcal{M}}^{0,k} y$  and  $y \preceq_{\mathcal{M}}^{0,k} z$ . This shows  $x \preceq_{\mathcal{M}}^{0,k} z$  and we conclude  $x \preceq_{\mathcal{M}}^{0,k} z$ . This proves that  $\preceq_{\mathcal{M}}^{0,k}$  is transitive.

Next we want to observe the following facts about  $\preceq_{\mathcal{M}}^{0,k}$ .

1. If  $x \preceq_{\mathcal{M}}^{0,k} y$  for some  $x, y \in A^*$  then  $x \preceq_{\mathcal{M}}^{0,k} wy$  for all  $w \in A^{\geq 2k+1}$ .
2. If  $x \preceq_{\mathcal{M}}^{0,k} y$  for some  $x, y \in A^*$  with  $\mathbf{p}_k(x) = \mathbf{p}_k(y)$  then  $x \preceq_{\mathcal{M}}^{0,k} y$ .

The first fact is easy to see from the definition of  $\preceq_{\mathcal{M}}^{0,k}$ . For the second one let  $x, y \in A^*$  such that  $x \preceq_{\mathcal{M}}^{0,k} y$  and  $\mathbf{p}_k(x) = \mathbf{p}_k(y)$ . Hence, there exist  $y_1 \in A^{\geq 2k+1} \cup \{\varepsilon\}$ ,  $y_2 \in A^*$  with  $y = y_1 y_2$  and  $x \preceq_{\mathcal{M}}^{0,k} y_2$ . If  $y_1 = \varepsilon$  then  $x \preceq_{\mathcal{M}}^{0,k} y$  and we are done. Otherwise we have  $|y_1| \geq 2k + 1$  and it follows that  $\mathbf{p}_k(x) = \mathbf{p}_k(y) = \mathbf{p}_k(y_1)$  and  $|x| \geq k$ . From  $x \preceq_{\mathcal{M}}^{0,k} y_2$  and Proposition 4.4 it follows that  $\mathbf{p}_k(x) = \mathbf{p}_k(y_2)$ . Therefore, with  $v =_{\text{def}} \mathbf{p}_k(x)$  we get  $|v| = k$ ,  $y_1 = v y'_1$  and  $y_2 = v y'_2$  for suitable  $y'_1, y'_2 \in A^*$ . It follows that  $|y'_1| \geq k + 1$  and we obtain  $y_2 = v y'_2 \prec_v^{0,k} v y'_1 v y'_2 = y_1 y_2 = y$ . Together with  $x \preceq_{\mathcal{M}}^{0,k} y_2$  this implies  $x \preceq_{\mathcal{M}}^{0,k} y$  which proves the second fact.

We turn to the proof of the theorem. By a length argument (namely that  $u \preceq_{\mathcal{M}}^{0,k} v$  with  $u \neq v$  implies  $|u| < |v|$ ) we only have to show that any set of pairwise incomparable elements is finite. Assume to the contrary that there is an infinite  $L \subseteq A^*$  such that all elements of  $L$  are pairwise incomparable with respect to  $\preceq_{\mathcal{M}}^{0,k}$ . Thus there is also an infinite  $L' \subseteq L$  such that all elements of  $L'$  have the same prefix of length  $k$ . By the second fact, elements of  $L'$  are pairwise incomparable w.r.t.  $\preceq_{\mathcal{M}}^{0,k}$ . In particular, this set  $L'$  witnesses the existence of infinite sequences  $\{f_i\}$  of words such that from  $i < j$  it follows that  $f_i \not\preceq_{\mathcal{M}}^{0,k} f_j$ . We will show that this is not true. For this consider any such sequence  $\{f_i\}$  and note that all words in such a sequence must be pairwise different since  $\preceq_{\mathcal{M}}^{0,k}$  is reflexive. We choose (using the axiom of choice) from all sequences  $\{f_i\}$  an ‘earliest’ sequence  $\{u_i\}$  as follows: let  $u_1$  be a shortest word beginning some sequence  $\{f_i\}$ , then let  $u_2$  be a shortest second word of any sequence  $u_1, f_2, f_3, \dots$ , then let  $u_3$  be a shortest third word of any sequence  $u_1, u_2, f_3, \dots$ , and so on. Clearly, also for  $\{u_i\}$  it holds that from  $i < j$  it follows that  $u_i \not\preceq_{\mathcal{M}}^{0,k} u_j$ . Since we have a finite alphabet there are words  $u_{i_1} = z w g_1, u_{i_2} = z w g_2, \dots$  with  $i_1 < i_2 < \dots$  for suitable  $z \in A^{2k+1}, w \in A^k, g_i \in A^+$ .

Now we look at the sequence  $u_1, u_2, \dots, u_{i_1-1}, w g_1, w g_2, \dots$  and denote it as  $\{x_i\}$ . Observe that this new sequence is ‘earlier’ than  $\{u_i\}$  since  $|w g_1| < |u_{i_1}|$ . In order to obtain a contradiction to our construction we need to show that for all  $i, j$  with  $i < j$  we can conclude  $x_i \preceq_{\mathcal{M}}^{0,k} x_j$ . This is clear if  $i, j \in \{1, \dots, i_1 - 1\}$  by the same property for  $\{u_i\}$ . Now

suppose  $i \in \{1, \dots, i_1 - 1\}$  and  $j \geq i_1$  and assume  $x_i \preceq_{\mathcal{M}}^{0,k} x_j$  where  $x_i = u_i$  and  $x_j = wg_l$  for some  $l \geq 1$ . By the first fact about  $\preceq_{\mathcal{M}}^{0,k}$  we have  $wg_l \preceq_{\mathcal{M}}^{0,k} zwg_l = u_{i_l}$  and together  $u_i \preceq_{\mathcal{M}}^{0,k} u_{i_l}$  (here we use the transitivity of  $\preceq_{\mathcal{M}}^{0,k}$ ), a contradiction. Finally let  $i, j \geq i_1$  with  $i < j$ . Assume  $x_i \preceq_{\mathcal{M}}^{0,k} x_j$  with  $x_i = wg_l$  and  $x_j = wg_m$  for some  $l < m$ . By our second fact about  $\preceq_{\mathcal{M}}^{0,k}$  we have  $wg_l \preceq_{\mathcal{M}}^{0,k} wg_m$  and with Proposition 4.4 we get  $zwg_l \preceq_{\mathcal{M}}^{0,k} zwg_m$ , i.e.,  $u_{i_l} \preceq_{\mathcal{M}}^{0,k} u_{i_m}$ . Since  $\preceq_{\mathcal{M}}^{0,k}$  is a refinement of  $\preceq_{\mathcal{M}}^{0,k}$  we conclude  $u_i \preceq_{\mathcal{M}}^{0,k} u_{i_m}$ , again a contradiction.  $\square$

Interestingly, it seems to be difficult to find a direct proof that uses only  $\preceq_{\mathcal{M}}^{0,k}$  but not  $\preceq_{\mathcal{M}}^{0,k}$ . In fact, this is the reason why we introduced the weakened relation  $\preceq_{\mathcal{M}}^{0,k}$ .

**Corollary 4.9.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $L \subseteq A^*$ . If  $L$  is a  $\preceq_{\mathcal{M}}^{0,k}$  co-ideal then it is even finitely generated (i.e.,  $L = \langle D \rangle_{\preceq_{\mathcal{M}}^{0,k}}$  for a finite  $D \subseteq A^+$ ).*

*Proof.* This is an immediate consequence of Theorem 4.8 and Proposition 1.9.  $\square$

#### 4.1.4 Languages from $\mathcal{FP}(\mathbb{P}_0^{\mathcal{L}})$ and $\mathcal{FP}(\mathbb{P}_0^{\mathcal{B}})$ are $\preceq_{\mathcal{M}}^{0,k}$ Co-Ideals

So far, in this section we showed (i) that the  $\preceq_{\mathcal{M}}^{0,k}$  upward closure of a nonempty word is in  $\mathcal{B}_{1/2}$  (respectively,  $\mathcal{L}_{1/2}$  for  $k = 0$ ) and (ii) that  $\preceq_{\mathcal{M}}^{0,k}$  co-ideals are finitely generated. Together this implies that  $\preceq_{\mathcal{M}}^{0,k}$  co-ideals are in  $\mathcal{B}_{1/2}$  (respectively,  $\mathcal{L}_{1/2}$  for  $k = 0$ ). Now we show that the forbidden-pattern classes  $\mathcal{FP}(\mathbb{P}_0^{\mathcal{B}})$  (respectively,  $\mathcal{FP}(\mathbb{P}_0^{\mathcal{L}})$ ) are  $\preceq_{\mathcal{M}}^{0,k}$  co-ideals (respectively,  $\preceq_{\mathcal{M}}^{0,0}$  co-ideals).

**Theorem 4.10.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA with  $L(\mathcal{M}) \in \mathcal{FP}(\mathbb{P}_0^{\mathcal{B}})$  and let  $k \geq \mathcal{I}_{\mathcal{M}}$ . Then  $L(\mathcal{M})$  is a  $\preceq_{\mathcal{M}}^{0,k}$  co-ideal.*

*Proof.* Let  $\mathcal{M}$  and  $k$  be as above and assume that  $L(\mathcal{M})$  is not a  $\preceq_{\mathcal{M}}^{0,k}$  co-ideal. By Definition 4.2 there exist words  $y, y' \in A^+$  and  $v' \in A^k$  such that  $y \in L(\mathcal{M})$ ,  $y' \notin L(\mathcal{M})$  and  $y <_{v'}^{0,k} y'$ . Hence we have  $y = y_1 v' y_2$  and  $y' = y_1 v' u' v' y_2$  for suitable words  $y_1, y_2 \in A^*$  and  $u' \in A^{*\geq k+1}$ . If we apply Corollary 1.17 to the word  $v'$  then we obtain words  $v, v_1, v_2 \in A^*$  with  $1 \leq |v| \leq k$ ,  $\delta^v = \delta^{vv}$  and  $v' = v_1 v v_2$ . With  $x =_{\text{def}} y_1 v_1 v$ ,  $w =_{\text{def}} v_2 u' v_1 v$  and  $z =_{\text{def}} v_2 y_2$  we obtain  $y = xz$  and  $y' = xwz$ . From  $\delta^v = \delta^{vv}$  it follows that both states  $s_1 =_{\text{def}} \delta(s_0, x)$  and  $s_2 =_{\text{def}} \delta(s_1, w)$  have  $v$ -loops. Therefore, with  $p =_{\text{def}} (v, w) \in \mathbb{P}_0^{\mathcal{B}} = \mathcal{B}$  we obtain  $s_1 \xrightarrow{p} s_2$ . Since  $s_0 \xrightarrow{x} s_1 \xrightarrow{z} +$  and  $s_2 \xrightarrow{z} -$  the DFA  $\mathcal{M}$  has a pattern from  $\mathbb{P}_0^{\mathcal{B}}$ . It follows that  $L(\mathcal{M}) \notin \mathcal{FP}(\mathbb{P}_0^{\mathcal{B}})$ . This is a contradiction to the assumption.  $\square$

**Theorem 4.11.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA with  $L(\mathcal{M}) \in \mathcal{FP}(\mathbb{P}_0^{\mathcal{L}})$ . Then  $L(\mathcal{M})$  is a  $\preceq_{\mathcal{M}}^{0,0}$  co-ideal.*

*Proof.* Let  $\mathcal{M}$  be as above and assume that  $L(\mathcal{M})$  is not a  $\preceq_{\mathcal{M}}^{0,0}$  co-ideal. By Definition 4.2 there exist words  $y, y' \in A^+$  such that  $y \in L(\mathcal{M})$ ,  $y' \notin L(\mathcal{M})$  and  $y <_{\varepsilon}^{0,k} y'$ . Hence we have  $y = y_1 y_2$  and  $y' = y_1 w y_2$  for suitable words  $y_1, y_2 \in A^*$  and  $w \in A^+$ . Let  $p =_{\text{def}} (\varepsilon, w) \in \mathbb{P}_0^{\mathcal{L}} = \mathcal{L}$ ,  $s_1 =_{\text{def}} \delta(s_0, y_1)$  and  $s_2 =_{\text{def}} \delta(s_1, w)$ . It is easy to see that  $s_1 \xrightarrow{p} s_2$ . Since  $s_0 \xrightarrow{y_1} s_1 \xrightarrow{y_2} +$  and  $s_2 \xrightarrow{y_2} -$  the DFA  $\mathcal{M}$  has a pattern from  $\mathbb{P}_0^{\mathcal{L}}$ . This shows  $L(\mathcal{M}) \notin \mathcal{FP}(\mathbb{P}_0^{\mathcal{L}})$  which is a contradiction.  $\square$

### 4.1.5 $\mathcal{L}_{1/2}$ and $\mathcal{B}_{1/2}$ are decidable

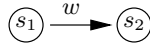
We combine the results of the preceding subsections and show the forbidden-pattern characterizations  $\mathcal{L}_{1/2} = \mathcal{FP}(\mathbb{P}_0^c)$  and  $\mathcal{B}_{1/2} = \mathcal{FP}(\mathbb{P}_0^s)$ . From the theory of forbidden-patterns in chapter 3 we obtain the decidability of  $\mathcal{L}_{1/2}$  and  $\mathcal{B}_{1/2}$ . These forbidden-pattern characterizations and their decidability conclusions were first shown in [Arf91, PW97] with an algebraic approach.

**Theorem 4.12.** *It holds that  $\mathcal{L}_{1/2} = \mathcal{FP}(\mathbb{P}_0^c)$  and  $\mathcal{B}_{1/2} = \mathcal{FP}(\mathbb{P}_0^s)$ .*

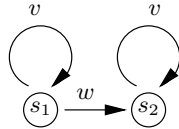
*Proof.* By Theorem 3.39 it suffices to show  $\mathcal{FP}(\mathbb{P}_0^c) \subseteq \mathcal{L}_{1/2}$  and  $\mathcal{FP}(\mathbb{P}_0^s) \subseteq \mathcal{B}_{1/2}$ . Let  $L \in \mathcal{FP}(\mathbb{P}_0^c)$ ,  $L' \in \mathcal{FP}(\mathbb{P}_0^s)$  and let  $\mathcal{M}, \mathcal{M}'$  be DFAs such that  $L = L(\mathcal{M})$  and  $L' = L(\mathcal{M}')$ . By the Theorems 4.10 and 4.11, it holds that (i)  $L$  is a  $\preceq_{\mathcal{M}}^{0,0}$  co-ideal and (ii)  $L'$  is a  $\preceq_{\mathcal{M}'}^{0,k}$  co-ideal for  $k =_{\text{def}} \mathcal{I}_{\mathcal{M}'}$ . From Corollary 4.9 it follows that there exist finite sets  $D, D' \subseteq A^+$  with

$$L = \langle D \rangle_{\preceq_{\mathcal{M}}^{0,0}} = \bigcup_{y \in D} \langle y \rangle_{\preceq_{\mathcal{M}}^{0,0}} \quad \text{and} \quad L' = \langle D' \rangle_{\preceq_{\mathcal{M}'}^{0,k}} = \bigcup_{y \in D'} \langle y \rangle_{\preceq_{\mathcal{M}'}^{0,k}}.$$

Finally, from Theorem 4.7 we get that  $\langle y \rangle_{\preceq_{\mathcal{M}}^{0,0}} \in \mathcal{L}_{1/2}$  and  $\langle y \rangle_{\preceq_{\mathcal{M}'}^{0,k}} \in \mathcal{B}_{1/2}$  for all  $y \in A^+$ . Since the unions above are finite, we obtain  $L \in \mathcal{L}_{1/2}$  and  $L' \in \mathcal{B}_{1/2}$ .  $\square$



**Fig. 4.2.** Forbidden-pattern for  $\mathcal{L}_{1/2}$  [Arf91, PW97] with  $w \in A^*$ .



**Fig. 4.3.** Forbidden-pattern for  $\mathcal{B}_{1/2}$  [PW97] with  $v, w \in A^+$ .

**Corollary 4.13.** *For every DFA  $\mathcal{M} = (A, S, \delta, s_0, S')$  the following holds.*

- $L(\mathcal{M}) \in \mathcal{L}_{1/2} \iff$  *there do not exist  $s_1, s_2 \in S$ ,  $z \in A^*$  such that  $s_0 \xrightarrow{z} s_1 \xrightarrow{z} +$ ,  $s_2 \xrightarrow{z} -$  and we find a pattern according to Figure 4.2 between  $s_1$  and  $s_2$ .*
- $L(\mathcal{M}) \in \mathcal{B}_{1/2} \iff$  *there do not exist  $s_1, s_2 \in S$ ,  $z \in A^*$  such that  $s_0 \xrightarrow{z} s_1 \xrightarrow{z} +$ ,  $s_2 \xrightarrow{z} -$  and we find a pattern according to Figure 4.3 between  $s_1$  and  $s_2$ .*

*Proof.* This follows from Theorem 4.12 and the definition of the forbidden-pattern classes (see Definitions 3.1–3.5).  $\square$

**Theorem 4.14.** *On input of a DFA  $\mathcal{M}$ , the questions  $L(\mathcal{M}) \in \mathcal{L}_{1/2}$  and  $L(\mathcal{M}) \in \mathcal{B}_{1/2}$  are decidable in nondeterministic logarithmic space.*

*Proof.* This is an immediate consequence of the Theorems 3.46 and 4.12.  $\square$

## 4.2 The Boolean Hierarchies over the Levels 1/2

A fundamental question came up recently in connection with complexity classes [BKS98]:

What is the minimal complexity of a given dot-depth one language in terms of Boolean combinations w.r.t.  $\mathcal{B}_{1/2}$ ?

In [BKS98] the authors define the Boolean hierarchy over  $\mathcal{B}_{1/2}$  in terms of classes  $\mathcal{B}_{1/2}(l)$  and  $\text{co}\mathcal{B}_{1/2}(l)$  for  $l \geq 1$ , and prove a levelwise correspondence to the Boolean hierarchy over NP via polynomial time leaf-languages. For further investigations in this direction an *effective* characterization of the single classes of the Boolean hierarchy over  $\mathcal{B}_{1/2}$  is desirable.

In this section we provide an effective characterization for all classes  $\mathcal{L}_{1/2}(l)$ ,  $\text{co}\mathcal{L}_{1/2}(l)$ ,  $\mathcal{B}_{1/2}(l)$  and  $\text{co}\mathcal{B}_{1/2}(l)$  of the Boolean hierarchies over  $\mathcal{L}_{1/2}$  and  $\mathcal{B}_{1/2}$ . A first proof for the decidability of the Boolean hierarchy over  $\mathcal{L}_{1/2}$  was given in [SW98] (using an automata-theoretic approach), a purely logical proof can be found in [Sel01].

We show even more, for a given language we can compute the exact level (i.e., the minimal level this languages belongs to) in the Boolean hierarchies over  $\mathcal{L}_{1/2}$  and  $\mathcal{B}_{1/2}$ . As a consequence, we can effectively compute an upper bound of the complexity class defined by a leaf-language from  $\mathcal{B}_1$  in the Boolean hierarchy over NP, which is related to the results in [BKS98].

We also show the strictness of the Boolean hierarchies over the levels 1/2 of the DDH and STH. More general strictness results concerning Boolean hierarchies over levels of concatenation hierarchies can be found in [Shu98, SS00]. There it is shown that the Boolean hierarchy over any level  $n + 1/2$  of the DDH and STH does not collapse. ([SS00] contains also a very interesting separability result: any two disjoint languages from  $\mathcal{L}_{3/2}$  are separable by a language from  $\mathcal{L}_{3/2} \cap \text{co}\mathcal{L}_{3/2}$ .)

In this section we use the technique of alternating chains. This technique was first used in model theory by Addison [Add65], and in recursion theory by Ershov [Ers68a, Ers68b]. Here we relate the Boolean level of a given language to the maximal number of alternations (w.r.t. to this language) in  $\preceq_{\mathcal{M}}^{0,k}$  chains.

**Definition 4.15.** *For a DFA  $\mathcal{M}$  and  $k \geq 0$  we define the maximal number of alternations in  $\preceq_{\mathcal{M}}^{0,k}$  chains as*

$$m_{\mathcal{M}}^{0,k} =_{\text{def}} \sup \left\{ n \geq 0 \mid \begin{array}{l} n = 0 \text{ or there exist words } w_0, \dots, w_n \in A^+ \text{ with } w_0 \in L(\mathcal{M}), \\ w_{i-1} \preceq_{\mathcal{M}}^{0,k} w_i \text{ and } w_{i-1} \in L(\mathcal{M}) \iff w_i \notin L(\mathcal{M}) \text{ for } 1 \leq i \leq n \end{array} \right\}.$$

### 4.2.1 $m_{\mathcal{M}}^{0,k}$ characterizes the Boolean Hierarchies over the Levels 1/2

Let  $\mathcal{M}$  be a DFA and choose  $k$  sufficiently large. In this subsection we show that  $L(\mathcal{M})$  is in level  $m_{\mathcal{M}}^{0,k} + 1$  but not in level  $m_{\mathcal{M}}^{0,k}$  of the Boolean hierarchy over  $\mathcal{B}_{1/2}$ . This means that with help of the measure  $m_{\mathcal{M}}^{0,k}$  we can determine the exact location of a given language in this hierarchy. We prove a similar result for the Boolean hierarchy over  $\mathcal{L}_{1/2}$ .

We start with an auxiliary result in Lemma 4.16 which is needed for the proof of Theorem 4.17. There we show that if  $k$  is large enough then for every  $\preceq_{\mathcal{M}}^{0,k}$  chain we find a  $\preceq_{\mathcal{M}}^{0,k+1}$  chain which has the same length and the same behavior w.r.t.  $\mathcal{M}$ . In particular this means that both chains have the same number of alternations w.r.t.  $L(\mathcal{M})$ . Then in Theorem 4.18 we show that if  $k$  is large enough then  $m_{\mathcal{M}}^{0,0} + 1$  and  $m_{\mathcal{M}}^{0,k} + 1$  tell us the levels of  $L(\mathcal{M})$  in the Boolean hierarchies over  $\mathcal{L}_{1/2}$  and  $\mathcal{B}_{1/2}$ , respectively. Finally, we give a strictness argument for these Boolean hierarchies.

**Lemma 4.16.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA and  $k \geq 3 \cdot \mathcal{I}_{\mathcal{M}}$ . Let  $n \geq 0$ ,  $y_0, \dots, y_n \in A^+$  and  $v_1, \dots, v_n \in A^k$  such that  $y_0 \prec_{v_1}^{0,k} y_1 \prec_{v_2}^{0,k} \dots \prec_{v_n}^{0,k} y_n$ . Then there exist a decomposition  $y_n = w_0 u_1 w_1 \dots u_m w_m$  and words  $y'_0, \dots, y'_n \in A^+$  such that:*

1.  $w_0, \dots, w_m, u_1, \dots, u_m \in A^{+\leq \mathcal{I}_{\mathcal{M}}}$  and  $\delta^{u_i} = \delta^{u_i u_i}$  for  $1 \leq i \leq m$
2.  $y'_0 \prec_{v'_1}^{0,k+1} y'_1 \prec_{v'_2}^{0,k+1} \dots \prec_{v'_n}^{0,k+1} y'_n$  for suitable words  $v'_1, \dots, v'_n \in A^{k+1}$
3.  $y'_n = w_0 u_1^{k+1} w_1 \dots u_m^{k+1} w_m$  and  $\delta^{y'_i} = \delta^{y_i}$  for  $0 \leq i \leq n$

*Proof.* We show the lemma by induction on  $n$ . For the induction base let  $n = 0$  and  $y_0 \in A^+$ . By Corollary 1.15 there exists a decomposition  $y_0 = w_0 u_1 w_1 \dots u_m w_m$  such that  $w_0, \dots, w_m, u_1, \dots, u_m \in A^{+\leq \mathcal{I}_{\mathcal{M}}}$  and  $\delta^{u_i} = \delta^{u_i u_i}$  for  $1 \leq i \leq m$ . Let  $y'_0 =_{\text{def}} w_0 u_1^{k+1} w_1 \dots u_m^{k+1} w_m$  and observe that  $\delta^{y'_0} = \delta^{y_0}$ . This shows the induction base.

We assume that there is some  $r \geq 0$  such that the lemma has been shown for  $n = r$  and we want to show it for  $n = r + 1$ . So let  $y_0, \dots, y_{r+1} \in A^+$  and  $v_1, \dots, v_{r+1} \in A^k$  such that  $y_0 \prec_{v_1}^{0,k} y_1 \prec_{v_2}^{0,k} \dots \prec_{v_{r+1}}^{0,k} y_{r+1}$ . By induction hypothesis there exist a decomposition  $y_r = w_0 u_1 w_1 \dots u_m w_m$  and words  $y'_0, \dots, y'_r \in A^+$  such that:

1.  $w_0, \dots, w_m, u_1, \dots, u_m \in A^{+\leq \mathcal{I}_{\mathcal{M}}}$  and  $\delta^{u_i} = \delta^{u_i u_i}$  for  $1 \leq i \leq m$
2.  $y'_0 \prec_{v'_1}^{0,k+1} y'_1 \prec_{v'_2}^{0,k+1} \dots \prec_{v'_r}^{0,k+1} y'_r$  for suitable words  $v'_1, \dots, v'_r \in A^{k+1}$
3.  $y'_r = w_0 u_1^{k+1} w_1 \dots u_m^{k+1} w_m$  and  $\delta^{y'_i} = \delta^{y_i}$  for  $0 \leq i \leq r$

Since  $y_r \prec_{v_{r+1}}^{0,k} y_{r+1}$  there exist  $x, z \in A^*$  and  $w \in A^{*\geq k+1}$  with  $y_r = x v_{r+1} z$  and  $y_{r+1} = x v_{r+1} w v_{r+1} z$ . Hence  $|y_r| \geq |v_{r+1}| = k \geq 3 \cdot \mathcal{I}_{\mathcal{M}}$  and it follows that  $m \geq 1$ . If we compare the decompositions  $y_r = x v_{r+1} z$  and  $y_r = w_0 u_1 w_1 \dots u_m w_m$  then (since  $|v_{r+1}| = k \geq 3 \cdot \mathcal{I}_{\mathcal{M}}$ ) there exists some  $1 \leq j \leq m$  such that  $u_j$  is a factor of  $v_{r+1}$ . This means that for suitable  $x', z' \in A^*$  it holds that  $v_{r+1} = x' u_j z'$ ,  $x x' = w_0 u_1 w_1 \dots u_{j-1} w_{j-1}$  and  $z' z = w_j u_{j+1} w_{j+1} \dots u_m w_m$ . Let  $\tilde{y} =_{\text{def}} z' w x'$  and observe that  $|\tilde{y}| \geq k + 2$  since  $|w| \geq k + 1$ ,  $|x' u_j z'| = |v_{r+1}| = k \geq 3 \cdot \mathcal{I}_{\mathcal{M}}$  and  $1 \leq |u_j| \leq \mathcal{I}_{\mathcal{M}}$ . From Corollary 1.15 we get a decomposition  $\tilde{y} = \tilde{w}_0 \tilde{u}_1 \tilde{w}_1 \dots \tilde{u}_{\tilde{m}} \tilde{w}_{\tilde{m}}$  such that  $\tilde{w}_0, \dots, \tilde{w}_{\tilde{m}}, \tilde{u}_1, \dots, \tilde{u}_{\tilde{m}} \in A^{+\leq \mathcal{I}_{\mathcal{M}}}$  and  $\delta^{\tilde{u}_i} = \delta^{\tilde{u}_i \tilde{u}_i}$  for  $1 \leq i \leq \tilde{m}$ . Therefore,  $y_{r+1}$  can be written as



$$\begin{aligned}
y_{r+1} &= x v_{r+1} w v_{r+1} z \\
&= \underbrace{xx'u_jz'} w \underbrace{x'u_jz'} z = xx'u_j \cdot \tilde{y} \cdot u_jz'z \\
&= w_0u_1w_1 \cdots u_{j-1}w_{j-1}u_j \cdot \underbrace{\tilde{w}_0\tilde{u}_1\tilde{w}_1 \cdots \tilde{u}_m\tilde{w}_m}_{\tilde{y}'=\text{def}} \cdot u_jw_ju_{j+1}w_{j+1} \cdots u_mw_m. \quad (4.3)
\end{aligned}$$

We will prove that (4.3) is the decomposition of  $y_{r+1}$  that is announced in the lemma. For this let  $y'_{r+1}$  be the word that emerges when we duplicate  $k+1$  times the factors  $u_i$  and  $\tilde{u}_i$  in  $y_{r+1}$ , i.e.,

$$y'_{r+1} =_{\text{def}} w_0u_1^{k+1}w_1 \cdots u_{j-1}^{k+1}w_{j-1}u_j^{k+1} \cdot \underbrace{\tilde{w}_0\tilde{u}_1^{k+1}\tilde{w}_1 \cdots \tilde{u}_m^{k+1}\tilde{w}_m}_{\tilde{y}'=\text{def}} \cdot u_j^{k+1}w_ju_{j+1}^{k+1}w_{j+1} \cdots u_m^{k+1}w_m.$$

Since for all  $i$  it holds that  $\delta^{u_i} = \delta^{u_i u_i}$  and  $\delta^{\tilde{u}_i} = \delta^{\tilde{u}_i \tilde{u}_i}$ , we get  $\delta^{y_{r+1}} = \delta^{y'_{r+1}}$ . So it remains to show  $y'_r <_{v'_{r+1}}^{0,k+1} y'_{r+1}$  for a suitable  $v'_{r+1} \in A^{k+1}$ .

From  $k+2 \leq |\tilde{y}'|$  it follows that  $\tilde{y}' \in A^{* \geq k+2}$ . Since  $|u_j| \geq 1$  we have  $u_j^{k+1} = \hat{v}v'_{r+1}$  for  $v'_{r+1} =_{\text{def}} \mathfrak{s}_{k+1}(u_j^{k+1})$  and a suitable  $\hat{v} \in A^*$ . Therefore,  $y'_r$  and  $y'_{r+1}$  can be written as

$$\begin{aligned}
y'_r &= w_0u_1^{k+1}w_1 \cdots u_{j-1}^{k+1}w_{j-1}\hat{v} \cdot v'_{r+1} \cdot w_ju_{j+1}^{k+1}w_{j+1} \cdots u_m^{k+1}w_m \quad \text{and} \\
y'_{r+1} &= w_0u_1^{k+1}w_1 \cdots u_{j-1}^{k+1}w_{j-1}\hat{v} \cdot v'_{r+1} \cdot \tilde{y}' \cdot \hat{v} \cdot v'_{r+1} \cdot w_ju_{j+1}^{k+1}w_{j+1} \cdots u_m^{k+1}w_m.
\end{aligned}$$

This shows  $y'_r <_{v'_{r+1}}^{0,k+1} y'_{r+1}$  and completes the induction step.  $\square$

**Theorem 4.17.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA,  $k \geq 3 \cdot \mathcal{I}_{\mathcal{M}}$ ,  $n \geq 0$  and  $y_0, \dots, y_n \in A^+$  such that  $y_0 \preceq_{\mathcal{M}}^{0,k} y_1 \preceq_{\mathcal{M}}^{0,k} \cdots \preceq_{\mathcal{M}}^{0,k} y_n$ . Then there exist words  $y'_0, \dots, y'_n \in A^+$  such that  $y'_0 \preceq_{\mathcal{M}}^{0,k+1} y'_1 \preceq_{\mathcal{M}}^{0,k+1} \cdots \preceq_{\mathcal{M}}^{0,k+1} y'_n$  and  $\delta^{y_i} = \delta^{y'_i}$  for  $0 \leq i \leq n$ .*

*Proof.* The definition of  $\preceq_{\mathcal{M}}^{0,k}$  implies that there exist an  $m \geq 0$ , words  $\hat{y}_0, \dots, \hat{y}_m \in A^+$ ,  $v_1, \dots, v_m \in A^k$  and indices  $0 = j_0 \leq j_1 \leq \cdots \leq j_m = m$  such that  $\hat{y}_0 <_{v_1}^{0,k} \hat{y}_1 <_{v_2}^{0,k} \cdots <_{v_m}^{0,k} \hat{y}_m$  and  $y_i = \hat{y}_{j_i}$  for  $0 \leq i \leq n$ . Lemma 4.16 implies in particular that there exist words  $\hat{y}'_0, \dots, \hat{y}'_m \in A^+$  and  $v'_1, \dots, v'_m \in A^{k+1}$  such that  $\hat{y}'_0 <_{v'_1}^{0,k+1} \hat{y}'_1 <_{v'_2}^{0,k+1} \cdots <_{v'_m}^{0,k+1} \hat{y}'_m$  and  $\delta^{\hat{y}_i} = \delta^{\hat{y}'_i}$  for  $0 \leq i \leq m$ . Therefore, if we define  $y'_i =_{\text{def}} \hat{y}'_{j_i}$  for  $0 \leq i \leq n$  then we obtain  $y'_0 \preceq_{\mathcal{M}}^{0,k+1} y'_1 \preceq_{\mathcal{M}}^{0,k+1} \cdots \preceq_{\mathcal{M}}^{0,k+1} y'_n$  and  $\delta^{y_i} = \delta^{y'_i}$  for  $0 \leq i \leq n$ .  $\square$

The main theorem of this subsection says that the measure  $m_{\mathcal{M}}^{0,k}$  characterizes the Boolean hierarchies over the levels 1/2 of the DDH and STH. This means that  $\mathcal{B}_{1/2}(n)$  is the class of languages that are accepted by a DFA  $\mathcal{M}$  with  $m_{\mathcal{M}}^{0,k} < n$  for a sufficiently large  $k$ . Analogously,  $\mathcal{L}_{1/2}(n)$  is the class of languages that are accepted by a DFA  $\mathcal{M}$  with  $m_{\mathcal{M}}^{0,0} < n$ .

**Theorem 4.18.** *Let  $\mathcal{M}$  be a DFA and  $k \geq 3 \cdot \mathcal{I}_{\mathcal{M}}$ . Then for all  $n \geq 1$  it holds that*

$$\begin{aligned}
m_{\mathcal{M}}^{0,k} < n &\iff L(\mathcal{M}) \in \mathcal{B}_{1/2}(n) \quad \text{and} \\
m_{\mathcal{M}}^{0,0} < n &\iff L(\mathcal{M}) \in \mathcal{L}_{1/2}(n).
\end{aligned}$$

*Proof.* We start with the implications from left to right and assume that  $m_{\mathcal{M}}^{0,k} < n$ . For  $m \geq 0$  let

$$L(m) =_{\text{def}} \left\{ w \in A^+ \left| \begin{array}{l} \text{there exist an } l \geq m \text{ and words } w_0, \dots, w_l \in A^+ \\ \text{such that } w_0 \in L(\mathcal{M}), w_l \preceq_{\mathcal{M}}^{0,k} w, w_{i-1} \preceq_{\mathcal{M}}^{0,k} w_i \text{ and} \\ w_{i-1} \in L(\mathcal{M}) \iff w_i \notin L(\mathcal{M}) \text{ for } 1 \leq i \leq l \end{array} \right. \right\}.$$

Hence it holds that  $L(0) \supseteq L(1) \supseteq L(2) \supseteq \dots$ . Moreover, from the definition of  $m_{\mathcal{M}}^{0,k}$  (i.e., Definition 4.15) it follows that  $L(m_{\mathcal{M}}^{0,k} + 1) = \emptyset$ . Let us observe that

$$L(\mathcal{M}) = L(0) \setminus (L(1) \setminus (L(2) \setminus \dots \setminus L(m_{\mathcal{M}}^{0,k}))) \dots. \quad (4.4)$$

If  $w \in L(\mathcal{M})$  then surely  $w \in L(0)$ . Therefore, there is some  $0 \leq j \leq m_{\mathcal{M}}^{0,k}$  such that  $w \in L(j) \setminus L(j+1)$ . Particularly it holds that  $w \in L(j) \setminus (L(j+1) \setminus (L(j+2) \setminus \dots \setminus L(m_{\mathcal{M}}^{0,k}))) \dots$ . Therefore, we obtain the following facts.

$$\begin{aligned} w &\in L(j) \setminus (L(j+1) \setminus (L(j+2) \setminus \dots \setminus L(m_{\mathcal{M}}^{0,k}))) \dots \\ w &\notin L(j-1) \setminus (L(j) \setminus (L(j+1) \setminus \dots \setminus L(m_{\mathcal{M}}^{0,k}))) \dots \\ w &\in L(j-2) \setminus (L(j-1) \setminus (L(j) \setminus \dots \setminus L(m_{\mathcal{M}}^{0,k}))) \dots \\ w &\notin L(j-3) \setminus (L(j-2) \setminus (L(j-1) \setminus \dots \setminus L(m_{\mathcal{M}}^{0,k}))) \dots \\ &\vdots \end{aligned}$$

It follows that  $w \in L(0) \setminus (L(1) \setminus (L(2) \setminus \dots \setminus L(m_{\mathcal{M}}^{0,k}))) \dots$  if and only if  $j \equiv 0 \pmod{2}$ . Since  $w \in L(j)$  there exist words  $w_0, \dots, w_j \in A^+$  such that  $w_0 \in L(\mathcal{M})$ ,  $w_j \preceq_{\mathcal{M}}^{0,k} w$ ,  $w_{i-1} \preceq_{\mathcal{M}}^{0,k} w_i$  and  $w_{i-1} \in L(\mathcal{M}) \iff w_i \notin L(\mathcal{M})$  for  $1 \leq i \leq j$ . Moreover, it holds that  $w_j \in L(\mathcal{M}) \iff w \in L(\mathcal{M})$  since otherwise we would obtain  $w \in L(j+1)$  by taking the word  $w_{j+1} =_{\text{def}} w$  into account. Hence  $w_j \in L(\mathcal{M})$  and it follows that  $j \equiv 0 \pmod{2}$ . From this we conclude  $w \in L(0) \setminus (L(1) \setminus (L(2) \setminus \dots \setminus L(m_{\mathcal{M}}^{0,k}))) \dots$ .

If  $w \in L(0) \setminus (L(1) \setminus (L(2) \setminus \dots \setminus L(m_{\mathcal{M}}^{0,k}))) \dots$  then we choose again some  $0 \leq j \leq m_{\mathcal{M}}^{0,k}$  such that  $w \in L(j) \setminus L(j+1)$ . As above we get words  $w_0, \dots, w_j \in A^+$  and it follows that

$$\begin{aligned} w_j \in L(\mathcal{M}) &\iff w \in L(\mathcal{M}) \quad \text{and} \\ w \in L(0) \setminus (L(1) \setminus (L(2) \setminus \dots \setminus L(m_{\mathcal{M}}^{0,k}))) \dots &\iff j \equiv 0 \pmod{2}. \end{aligned}$$

It follows that  $j \equiv 0 \pmod{2}$ . Therefore,  $w_j \in L(\mathcal{M})$  and we conclude  $w \in L(\mathcal{M})$ . This shows equation (4.4).

By definition,  $L(i)$  is a  $\preceq_{\mathcal{M}}^{0,k}$  co-ideal for all  $i \geq 0$ . Together with  $L(i) \subseteq A^+$  and Corollary 4.9 this implies that for all  $i \geq 0$  it holds that  $L(i) = \langle D \rangle_{\preceq_{\mathcal{M}}^{0,k}}$  for some finite set  $D \subseteq A^+$ . Now Theorem 4.7 shows that  $L(i) \in \mathcal{B}_{1/2}$  for  $i \geq 0$ . Since  $m_{\mathcal{M}}^{0,k} < n$  it follows that  $L(\mathcal{M}) = L(0) \setminus (L(1) \setminus (L(2) \setminus \dots \setminus L(m_{\mathcal{M}}^{0,k}))) \dots \in \mathcal{B}_{1/2}(m_{\mathcal{M}}^{0,k} + 1) \subseteq \mathcal{B}_{1/2}(n)$ . Analogously we obtain that  $L(\mathcal{M}) \in \mathcal{L}_{1/2}(n)$  is implied by  $m_{\mathcal{M}}^{0,0} < n$ .

We turn to the proof of the implications from right to left. For this we assume that  $L(\mathcal{M}) \in \mathcal{B}_{1/2}(n)$  and  $m_{\mathcal{M}}^{0,k} \geq n$ . This will lead to a contradiction. Our assumption implies  $L(\mathcal{M}) = L_0 \setminus (L_1 \setminus (L_2 \setminus \dots \setminus L(n-1))) \dots$  for suitable languages  $L_0, \dots, L_{n-1} \in \mathcal{B}_{1/2}$

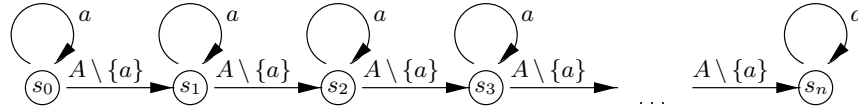
with  $L_0 \supseteq L_1 \supseteq \dots \supseteq L_{n-1}$ . For  $0 \leq i \leq n-1$  let  $\mathcal{M}_i$  be a DFA with  $L_i = L(\mathcal{M}_i)$  and let  $k' =_{\text{def}} \max(\{3 \cdot \mathcal{I}_{\mathcal{M}_i} \mid 0 \leq i \leq n-1\} \cup \{k\})$ . Since  $m_{\mathcal{M}}^{0,k} \geq n \geq 1$  there exist words  $w_0, \dots, w_n \in A^+$  with  $w_0 \in L(\mathcal{M})$  and  $w_0 \preceq_{\mathcal{M}}^{0,k} w_1 \preceq_{\mathcal{M}}^{0,k} \dots \preceq_{\mathcal{M}}^{0,k} w_n$  such that  $w_{i-1} \in L(\mathcal{M}) \iff w_i \notin L(\mathcal{M})$  for  $1 \leq i \leq n$ . If we apply Theorem 4.17 repeatedly to the chain  $w_0 \preceq_{\mathcal{M}}^{0,k} w_1 \preceq_{\mathcal{M}}^{0,k} \dots \preceq_{\mathcal{M}}^{0,k} w_n$  then we obtain a chain  $w'_0 \preceq_{\mathcal{M}}^{0,k'} w'_1 \preceq_{\mathcal{M}}^{0,k'} \dots \preceq_{\mathcal{M}}^{0,k'} w'_n$  of nonempty words such that  $w'_0 \in L(\mathcal{M})$  and  $w'_{i-1} \in L(\mathcal{M}) \iff w'_i \notin L(\mathcal{M})$  for  $1 \leq i \leq n$ .

By Theorem 4.12 we have  $L(\mathcal{M}_i) \in \mathcal{FP}(\mathbb{P}_0^{\mathcal{B}})$  for  $0 \leq i \leq n-1$ . With Theorem 4.10 and  $k' \geq \mathcal{I}_{\mathcal{M}_i}$  we obtain that  $L(\mathcal{M}_i)$  is a  $\preceq_{\mathcal{M}}^{0,k'}$  co-ideal for  $0 \leq i \leq n-1$ . Moreover, from  $w'_{j-1} \preceq_{\mathcal{M}}^{0,k'} w'_j$  for  $1 \leq j \leq n$  and Proposition 4.3 it follows that  $w'_{j-1} \preceq_{\mathcal{M}_i}^{0,k'} w'_j$  for  $1 \leq j \leq n$  and  $0 \leq i \leq n-1$ . Therefore, it holds that

$$w'_{j-1} \in L_i \implies w'_j \in L_i \quad \text{for } 1 \leq j \leq n \text{ and } 0 \leq i \leq n-1. \quad (4.5)$$

Since  $L_0 \supseteq L_1 \supseteq \dots \supseteq L_{n-1}$  and  $w'_0 \in L(\mathcal{M})$  we have  $w'_0 \in L_0$ . From (4.5) we get  $w'_1 \in L_0$ . If  $w'_0 \in L_1$  then also  $w'_1 \in L_1$  by (4.5). If  $w'_0 \notin L_1$  then it follows that  $w'_1 \in L_1$  (otherwise we would have  $w'_0, w'_1 \in L_0 \setminus L_1$  which contradicts the assumption that  $w'_0 \in L(\mathcal{M}) \iff w'_1 \notin L(\mathcal{M})$ ). So in both cases we obtain  $w'_1 \in L_1$ , and analogously we get  $w'_2 \in L_2, w'_3 \in L_3, \dots, w'_{n-1} \in L_{n-1}$ .

From  $w'_{n-1} \in L_{n-1}$  and (4.5) we get  $w'_n \in L_{n-1}$ . So  $w'_{n-1} \in L(\mathcal{M}) \iff w'_n \in L(\mathcal{M})$  which is a contradiction. We conclude that  $L(\mathcal{M}) \in \mathcal{B}_{1/2}(n)$  implies  $m_{\mathcal{M}}^{0,k} < n$ . Analogously we obtain  $L(\mathcal{M}) \in \mathcal{L}_{1/2}(n) \implies m_{\mathcal{M}}^{0,0} < n$ .  $\square$



**Fig. 4.4.** Definition of the DFA  $\mathcal{M}_n$  where  $n \geq 0$ ,  $a \in A$  and the state  $s_i$  is accepting if and only if  $i \equiv 0 \pmod{2}$ .

Next, we show the strictness of the Boolean hierarchies over the levels 1/2 of the DDH and STH. These are known results from [Shu98, SS00]. There it is shown that both Boolean hierarchies over  $\mathcal{L}_{n+1/2}$  and over  $\mathcal{B}_{n+1/2}$  are strict for all  $n \geq 0$ .

**Corollary 4.19.** *The Boolean hierarchies over  $\mathcal{L}_{1/2}$  and  $\mathcal{B}_{1/2}$  are strict.*

*Proof.* By Theorem 4.18 it suffices to show that for each  $n \geq 0$  there exist a DFA  $\mathcal{M}$  such that  $m_{\mathcal{M}}^{0,k} = n$  for all  $k \geq 0$ . Fix some  $n \geq 0$ , choose different letters  $a, b \in A$  and let  $\mathcal{M} =_{\text{def}} \mathcal{M}_n$  be the DFA in Figure 4.4.

Let  $k \geq 0$  and  $w_i =_{\text{def}} a^{2k}(ba^{2k})^i$  for  $0 \leq i \leq n$ . It follows that  $w_i \preceq_{\mathcal{M}}^{0,k} w_{i+1}$  for  $0 \leq i < n$ . Moreover, from the definition of  $\mathcal{M}$  in Figure 4.4 we get  $w_i \in L(\mathcal{M}) \iff i \equiv 0 \pmod{2}$  for  $0 \leq i \leq n$ . This shows  $m_{\mathcal{M}}^{0,k} \geq n$ .

Suppose  $m_{\mathcal{M}}^{0,k} > n$ . This means that there exist an  $m > n$  and words  $y_0, \dots, y_m \in A^+$  with  $y_0 \preceq_{\mathcal{M}}^{0,k} y_1 \preceq_{\mathcal{M}}^{0,k} \dots \preceq_{\mathcal{M}}^{0,k} y_m$  such that for  $0 \leq i \leq m$  it holds that  $y_i \in L(\mathcal{M})$  if and only if  $i \equiv 0 \pmod{2}$ .

Note that each state in  $\mathcal{M}$  has an  $a$ -loop. Therefore, if we insert letters  $a$  into a given word then the emerging word has the same acceptance behavior w.r.t.  $\mathcal{M}$ . So if we want to change the acceptance behavior then we have to insert at least one letter from  $A \setminus \{a\}$ . It follows that both words  $y_{m-1}$  and  $y_m$  contain at least  $m-1 \geq n$  letters from  $A \setminus \{a\}$ . Now we look at Figure 4.4 again and we see that  $\delta(s_0, y_{m-1}) = s_n = \delta(s_0, y_m)$ . This contradicts the fact that  $y_i \in L(\mathcal{M})$  if and only if  $i \equiv 0 \pmod{2}$  for  $0 \leq i \leq m$ . So we conclude that  $m_{\mathcal{M}}^{0,k} = n$ .  $\square$

Note that we showed more than the lemma: The Boolean hierarchies over  $\mathcal{L}_{1/2}$  and  $\mathcal{B}_{1/2}$  can be separated by the same family of languages. In particular,  $L(\mathcal{M}_n)$  is in  $\mathcal{L}_{1/2}(n+1)$  but not in  $\mathcal{B}_{1/2}(n)$  for  $n \geq 1$ .

#### 4.2.2 $m_{\mathcal{M}}^{0,k}$ is computable

In this subsection we show that the Boolean hierarchies over  $\mathcal{L}_{1/2}$  and over  $\mathcal{B}_{1/2}$  are decidable. From the previous subsection we know that it suffices to show that the question  $m_{\mathcal{M}}^{0,k} < n$  is decidable.

We start this subsection with two decomposition lemmas for  $\preceq_{\mathcal{M}}^{0,k}$  word extensions. In both lemmas we consider different words  $w, w'$  with  $w \preceq_{\mathcal{M}}^{0,k} w'$ . The first lemma finds a context word  $v$  and decompositions  $w = w_1 v w_2$  and  $w' = w'_1 v w'_2$  such that  $w_1 v \preceq_{\mathcal{M}}^{0,k} w'_1 v$  and  $v w_2 \preceq_{\mathcal{M}}^{0,k} v w'_2$ . This means that any extension  $w \preceq_{\mathcal{M}}^{0,k} w'$  can be divided into the following independent parts: (i)  $v \prec_v^{0,k} v w v$ , (ii)  $w_1 v \preceq_{\mathcal{M}}^{0,k} w'_1 v$  and (iii)  $v w_2 \preceq_{\mathcal{M}}^{0,k} v w'_2$ . The second lemma strengthens the first one. There we prove that unless the extension  $w \preceq_{\mathcal{M}}^{0,k} w'$  has a very simple structure we can even find *nonempty* words  $w_1$  and  $w_2$ .

Then in Theorem 4.26 we show that the maximal number of alternations in  $\preceq_{\mathcal{M}}^{0,k}$  chains already appears in  $\preceq_{\mathcal{M}}^{0,k}$  chains that contain only short words. This allows to decide the question  $m_{\mathcal{M}}^{0,k} < n$ , and therefore this yields the decidability of the Boolean hierarchies over the levels 1/2. Moreover, for a given language we can compute the exact level in these Boolean hierarchies (i.e., the minimal level this languages belongs to).

**Lemma 4.20.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $w, w' \in A^*$  with  $w \preceq_{\mathcal{M}}^{0,k} w'$  and  $w \neq w'$ . Then there exist words  $w_1, w_2, w'_1, w'_2 \in A^*$ ,  $v \in W_{\mathcal{M}}^{0,k}$  and  $u \in A^{*\geq k+1}$  such that  $w = w_1 v w_2$ ,  $w' = w'_1 v w'_2$ ,  $w_1 v \preceq_{\mathcal{M}}^{0,k} w'_1 v$  and  $v w_2 \preceq_{\mathcal{M}}^{0,k} v w'_2$ .*

*Proof.* Let  $\mathcal{M}$  be a DFA and  $k \geq 0$ . By the definition of  $\preceq_{\mathcal{M}}^{0,k}$  it suffices to show the lemma for chains of  $\prec_v^{0,k}$  extensions. Therefore, it is enough to prove the following claim.

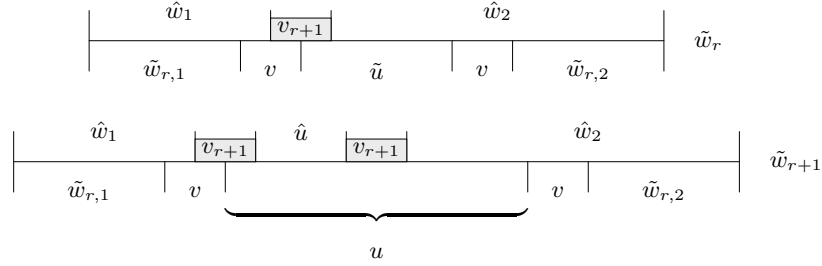
**Claim.** *Let  $n \geq 1$ ,  $w, w', \tilde{w}_0, \dots, \tilde{w}_n \in A^*$  and  $v_1, \dots, v_n \in W_{\mathcal{M}}^{0,k}$  with  $w = \tilde{w}_0 \prec_{v_1}^{0,k} \tilde{w}_1 \prec_{v_2}^{0,k} \dots \prec_{v_n}^{0,k} \tilde{w}_n = w'$ . Then there exist words  $w_1, w_2, w'_1, w'_2 \in A^*$ ,  $v \in W_{\mathcal{M}}^{0,k}$  and  $u \in A^{*\geq k+1}$  such that  $w = w_1 v w_2$ ,  $w' = w'_1 v w'_2$ ,  $w_1 v \preceq_{\mathcal{M}}^{0,k} w'_1 v$  and  $v w_2 \preceq_{\mathcal{M}}^{0,k} v w'_2$ .*

We prove the claim by induction on  $n \geq 1$ . For  $n = 1$  we have  $w \prec_{v_1}^{0,k} w'$ . Hence there exist words  $w_1, w_2 \in A^*$  and  $u \in A^{*\geq k+1}$  such that  $w = w_1 v_1 w_2$  and  $w' = w_1 v_1 u v_1 w_2$ . The induction base follows with  $w'_1 =_{\text{def}} w_1$ ,  $w'_2 =_{\text{def}} w_2$  and  $v =_{\text{def}} v_1$ .

Assume that there is some  $r \geq 1$  such that the claim has been shown for all  $n \leq r$ . Now we consider  $w, w', \tilde{w}_0, \dots, \tilde{w}_{r+1} \in A^*$  and  $v_1, \dots, v_{r+1} \in W_{\mathcal{M}}^{0,k}$  with  $w = \tilde{w}_0 \prec_{v_1}^{0,k} \tilde{w}_1 \prec_{v_2}^{0,k}$

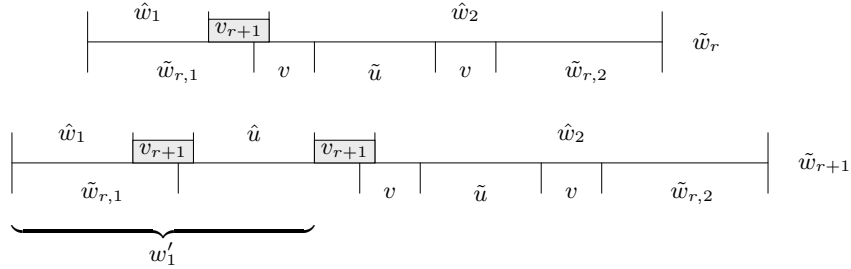
$\dots \prec_{v_{r+1}}^{0,k} \tilde{w}_{r+1} = w'$ . By induction hypothesis, there exist words  $w_1, w_2, \tilde{w}_{r,1}, \tilde{w}_{r,2} \in A^*$ ,  $v \in W_{\mathcal{M}}^{0,k}$  and  $\tilde{u} \in A^{*\geq k+1}$  such that  $w = w_1 v w_2$ ,  $\tilde{w}_r = \tilde{w}_{r,1} v \tilde{u} v \tilde{w}_{r,2}$ ,  $w_1 v \preceq_{\mathcal{M}}^{0,k} \tilde{w}_{r,1} v$  and  $v w_2 \preceq_{\mathcal{M}}^{0,k} v \tilde{w}_{r,2}$ . Moreover, since  $\tilde{w}_r \prec_{v_{r+1}}^{0,k} \tilde{w}_{r+1}$  there exist words  $\hat{w}_1, \hat{w}_2 \in A^*$  and  $\hat{u} \in A^{*\geq k+1}$  such that  $\tilde{w}_r = \hat{w}_1 v_{r+1} \hat{w}_2$  and  $\tilde{w}_{r+1} = \hat{w}_1 v_{r+1} \hat{u} v_{r+1} \hat{w}_2$ . In the following we compare the decompositions  $\tilde{w}_r = \tilde{w}_{r,1} v \tilde{u} v \tilde{w}_{r,2}$  and  $\tilde{w}_r = \hat{w}_1 v_{r+1} \hat{w}_2$ .

**Case 1:** Assume that the factor  $v_{r+1}$  of the decomposition  $\tilde{w}_r = \hat{w}_1 v_{r+1} \hat{w}_2$  is contained in the factor  $v \tilde{u} v$  of the decomposition  $\tilde{w}_r = \tilde{w}_{r,1} v \tilde{u} v \tilde{w}_{r,2}$ .



It follows that  $\tilde{w}_{r+1} \in \tilde{w}_{r,1} \cdot \langle v \tilde{u} v \rangle_{\prec_{v_{r+1}}^{0,k}} \cdot \tilde{w}_{r,2}$ . Since  $\langle v \tilde{u} v \rangle_{\prec_{v_{r+1}}^{0,k}} \subseteq v A^{*\geq k+1} v$  there exists some  $u \in A^{*\geq k+1}$  such that  $\tilde{w}_{r+1} = \tilde{w}_{r,1} v u v \tilde{w}_{r,2}$ . With  $w'_1 =_{\text{def}} \tilde{w}_{r,1}$  and  $w'_2 =_{\text{def}} \tilde{w}_{r,2}$  we get  $w = w_1 v w_2$ ,  $w' = w'_1 v u v w'_2$ ,  $w_1 v \preceq_{\mathcal{M}}^{0,k} w'_1 v$  and  $v w_2 \preceq_{\mathcal{M}}^{0,k} v w'_2$ .

**Case 2:** Assume that the factor  $v_{r+1}$  of the decomposition  $\tilde{w}_r = \hat{w}_1 v_{r+1} \hat{w}_2$  is *not* contained in the factor  $v \tilde{u} v$  of the decomposition  $\tilde{w}_r = \tilde{w}_{r,1} v \tilde{u} v \tilde{w}_{r,2}$ . So either  $v_{r+1}$  is a factor of  $\tilde{w}_{r,1} v$  or is a factor of  $v \tilde{w}_{r,2}$ . Without loss of generality we assume the former.



Hence,  $\tilde{w}_{r+1} \in \langle \tilde{w}_{r,1} v \rangle_{\prec_{v_{r+1}}^{0,k}} \cdot \tilde{u} v \tilde{w}_{r,2}$ . Since  $\langle \tilde{w}_{r,1} v \rangle_{\prec_{v_{r+1}}^{0,k}} \subseteq A^* v$  there exists a word  $w'_1 \in A^*$  with  $\tilde{w}_{r+1} = w'_1 v \cdot \tilde{u} v \tilde{w}_{r,2}$  and  $\tilde{w}_{r,1} v \preceq_{\mathcal{M}}^{0,k} w'_1 v$ . With  $w'_2 =_{\text{def}} \tilde{w}_{r,2}$  and  $u =_{\text{def}} \tilde{u}$  we get  $w = w_1 v w_2$ ,  $w' = w'_1 v u v w'_2$ ,  $w_1 v \preceq_{\mathcal{M}}^{0,k} w'_1 v$  and  $v w_2 \preceq_{\mathcal{M}}^{0,k} v w'_2$ . This completes the induction step and the claim follows.  $\square$

With the following lemma we strengthen Lemma 4.20: We may assume that the words  $w_1$  and  $w_2$  are nonempty unless the extension  $w \preceq_{\mathcal{M}}^{0,k} w'$  has a very simple structure.

**Lemma 4.21.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $w, w' \in A^*$  with  $w \preceq_{\mathcal{M}}^{0,k} w'$ . If  $w'$  does not belong to  $(\mathfrak{p}_k(w)A^{*\geq k+1} \cup \{\varepsilon\}) \cdot w \cdot (A^{*\geq k+1}\mathfrak{s}_k(w) \cup \{\varepsilon\})$  then there exist words  $w_1, w_2, w'_1, w'_2 \in A^+$ ,  $v \in W_{\mathcal{M}}^{0,k}$  and  $u \in A^{*\geq k+1}$  with  $w = w_1vw_2$ ,  $w' = w'_1vuw'_2$ ,  $w_1v \preceq_{\mathcal{M}}^{0,k} w'_1v$  and  $vw_2 \preceq_{\mathcal{M}}^{0,k} vw'_2$ .*

*Proof.* We show the lemma by induction on  $|w'| \geq 0$ . First of all it is easy to see that if  $w' \notin (\mathfrak{p}_k(w)A^{*\geq k+1} \cup \{\varepsilon\}) \cdot w \cdot (A^{*\geq k+1}\mathfrak{s}_k(w) \cup \{\varepsilon\})$  then  $w' \neq w$ . For the induction base we have to consider words  $w, w' \in A^*$  with  $w \preceq_{\mathcal{M}}^{0,k} w'$  and  $w' = \varepsilon$ . This implies  $w = w'$  which shows the induction base.

Assume that there is some  $r \geq 0$  such that the lemma has been shown for all  $w, w' \in A^*$  with  $|w'| \leq r$ . Now let  $w, w' \in A^*$  with  $w \preceq_{\mathcal{M}}^{0,k} w'$  and  $|w'| = r + 1$ . If  $w = w'$  then we are done. So from now on we assume that  $w \neq w'$ .

From Lemma 4.20 we get words  $w_1, w_2, w'_1, w'_2 \in A^*$ ,  $v \in W_{\mathcal{M}}^{0,k}$  and  $u \in A^{*\geq k+1}$  with  $w = w_1vw_2$ ,  $w' = w'_1vuw'_2$ ,  $w_1v \preceq_{\mathcal{M}}^{0,k} w'_1v$  and  $vw_2 \preceq_{\mathcal{M}}^{0,k} vw'_2$ . If  $w_1, w_2 \in A^+$  then it follows that also  $w'_1, w'_2 \in A^+$  and we are done. From now on we assume that  $w_1 = \varepsilon$  (the case where we start with the assumption  $w_2 = \varepsilon$  can be shown analogously).

If also  $w_2 = \varepsilon$  then  $w = v$ . From  $w \preceq_{\mathcal{M}}^{0,k} w'$  and  $w \neq w'$  it follows that  $w' \in wA^{*\geq k+1}w \subseteq (\mathfrak{p}_k(w)A^{*\geq k+1} \cup \{\varepsilon\}) \cdot w \cdot (A^{*\geq k+1}\mathfrak{s}_k(w) \cup \{\varepsilon\})$ . So if  $w_1 = w_2 = \varepsilon$  then we are done. Hence from now on we assume  $w_2 \neq \varepsilon$ .

Next we want to see that we may also assume  $w'_1 = \varepsilon$ . Suppose that  $w'_1 \in A^+$ . By Proposition 4.4,  $w'$  has  $v$  as a prefix since  $vw_2 = w \preceq_{\mathcal{M}}^{0,k} w'$ . Therefore, if we define  $\tilde{u} =_{\text{def}} A^{-k}(w'_1vu)$  and if we use  $\varepsilon$  instead of  $w'_1$  we obtain  $\tilde{u} \in A^{*\geq k+1}$ ,  $w = \varepsilon vw_2$ ,  $w' = \varepsilon v\tilde{u}w'_2$ ,  $\varepsilon v \preceq_{\mathcal{M}}^{0,k} \varepsilon v$  and  $vw_2 \preceq_{\mathcal{M}}^{0,k} vw'_2$ . Hence, from now on we assume  $w'_1 = \varepsilon$ . Moreover, since  $vw_2 \preceq_{\mathcal{M}}^{0,k} vw'_2$  and  $w_2 \in A^+$  it follows that  $w'_2 \neq \varepsilon$ . So we have reached the following situation:  $w_1 = w'_1 = \varepsilon$  and  $w_2, w'_2 \in A^+$ .

Let  $\tilde{w} =_{\text{def}} vw_2$  and  $\tilde{w}' =_{\text{def}} vw'_2$ . It follows that  $\tilde{w} \preceq_{\mathcal{M}}^{0,k} \tilde{w}'$  and  $|v|+1 \leq |\tilde{w}| \leq |\tilde{w}'| < |w'|$ . A comparison of the words  $\tilde{w}$  and  $\tilde{w}'$  leads to two cases.

**Case 1:** Assume that  $\tilde{w}' \in (\mathfrak{p}_k(\tilde{w})A^{*\geq k+1} \cup \{\varepsilon\}) \cdot \tilde{w} \cdot (A^{*\geq k+1}\mathfrak{s}_k(\tilde{w}) \cup \{\varepsilon\})$ . From  $\tilde{w} = w$  it follows that  $\tilde{w}' \in (\mathfrak{p}_k(w)A^{*\geq k+1} \cup \{\varepsilon\}) \cdot w \cdot (A^{*\geq k+1}\mathfrak{s}_k(w) \cup \{\varepsilon\})$ . Therefore, we obtain  $w' = v\tilde{w}' = \mathfrak{p}_k(w)u\tilde{w}' \in (\mathfrak{p}_k(w)A^{*\geq k+1} \cup \{\varepsilon\}) \cdot w \cdot (A^{*\geq k+1}\mathfrak{s}_k(w) \cup \{\varepsilon\})$  and we are done.

**Case 2:** Assume that  $\tilde{w}' \notin (\mathfrak{p}_k(\tilde{w})A^{*\geq k+1} \cup \{\varepsilon\}) \cdot \tilde{w} \cdot (A^{*\geq k+1}\mathfrak{s}_k(\tilde{w}) \cup \{\varepsilon\})$ . Since  $|\tilde{w}'| \leq r$  we can apply the induction hypothesis to  $\tilde{w} \preceq_{\mathcal{M}}^{0,k} \tilde{w}'$ . We obtain words  $\tilde{w}_1, \tilde{w}_2, \tilde{w}'_1, \tilde{w}'_2 \in A^+$ ,  $\tilde{v} \in W_{\mathcal{M}}^{0,k}$  and  $\tilde{u} \in A^{*\geq k+1}$  with  $\tilde{w} = \tilde{w}_1\tilde{v}\tilde{w}_2$ ,  $\tilde{w}' = \tilde{w}'_1\tilde{v}\tilde{u}\tilde{w}'_2$ ,  $\tilde{w}_1\tilde{v} \preceq_{\mathcal{M}}^{0,k} \tilde{w}'_1\tilde{v}$  and  $\tilde{v}\tilde{w}_2 \preceq_{\mathcal{M}}^{0,k} \tilde{v}\tilde{w}'_2$ . Let  $\hat{w}_1 =_{\text{def}} \tilde{w}_1$ ,  $\hat{w}_2 =_{\text{def}} \tilde{w}_2$ ,  $\hat{w}'_1 =_{\text{def}} v\tilde{u}\tilde{w}'_1$  and  $\hat{w}'_2 =_{\text{def}} \tilde{w}'_2$ . Then it holds that  $\hat{w}_1, \hat{w}_2, \hat{w}'_1, \hat{w}'_2 \in A^+$ . Note that  $\mathfrak{p}_k(\tilde{w}') = v$  and that  $\tilde{w}'_1\tilde{v}$  is a prefix of  $\tilde{w}'$  with length  $\geq |\tilde{v}| = k$ . It follows that  $v$  is also a prefix of  $\tilde{w}'_1\tilde{v}$ . Hence  $\tilde{w}'_1\tilde{v} <_{\mathcal{M}}^{0,k} v\tilde{u}\tilde{w}'_1\tilde{v} = \hat{w}'_1\tilde{v}$  and we obtain

$$\begin{aligned} w &= \tilde{w} = \hat{w}_1\tilde{v}\hat{w}_2, \\ w' &= v\tilde{u}\tilde{w}'_2 = v\tilde{u}\tilde{w}'_1\tilde{v}\tilde{u}\tilde{w}'_2 = \hat{w}'_1\tilde{v}\tilde{u}\tilde{w}'_2, \\ \hat{w}_1\tilde{v} &= \tilde{w}_1\tilde{v} \preceq_{\mathcal{M}}^{0,k} \tilde{w}'_1\tilde{v} \preceq_{\mathcal{M}}^{0,k} \hat{w}'_1\tilde{v} \quad \text{and} \\ \tilde{v}\hat{w}_2 &= \tilde{v}\tilde{w}_2 \preceq_{\mathcal{M}}^{0,k} \tilde{v}\tilde{w}'_2 = \tilde{v}\hat{w}'_2. \end{aligned}$$

This completes the induction step.  $\square$

In order to bound the lengths of words in alternating  $\preceq_{\mathcal{M}}^{0,k}$  chains we define below two bounding functions. Then in Lemma 4.24 we will show that every extension  $w \preceq_{\mathcal{M}}^{0,k} w'$  where  $w'$  is substantial longer than  $w$  can be written as  $w \preceq_{\mathcal{M}}^{0,k} \hat{w} \preceq_{\mathcal{M}}^{0,k} w'$  such that  $\hat{w}$  is a reasonable short word equivalent to  $w'$ . We will use this lemma to obtain Theorem 4.26 which shows that the maximal number of alternations in  $\preceq_{\mathcal{M}}^{0,k}$  chains already appears in  $\preceq_{\mathcal{M}}^{0,k}$  chains containing only short words. Together with the characterizations of the Boolean hierarchies from subsection 4.2.1 this yields the decidability of the Boolean hierarchies over the levels 1/2.

**Definition 4.22.** Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA and  $n, k \geq 0$ .

$$\mathcal{D}_{\mathcal{M}}^k(n) \stackrel{\text{def}}{=} \begin{cases} 2 \cdot (2k+1) \cdot (|A|^k \cdot |\mathcal{M}|^{|\mathcal{M}|} + 2) & : \text{ if } n = 0 \\ 3 \cdot \mathcal{D}_{\mathcal{M}}^k(n-1) + n & : \text{ otherwise} \end{cases}$$

$$\mathcal{E}_{\mathcal{M}}^k(n) \stackrel{\text{def}}{=} \begin{cases} \mathcal{D}_{\mathcal{M}}^k(0) & : \text{ if } n = 0 \\ \mathcal{D}_{\mathcal{M}}^k(\mathcal{E}_{\mathcal{M}}^k(n-1)) & : \text{ otherwise} \end{cases}$$

From this definition it is easy to see that  $\mathcal{D}_{\mathcal{M}}^k(\cdot)$  is a monotone increasing function such that the following holds for every DFA  $\mathcal{M}$  and all  $k, n \geq 0$ .

$$\mathcal{D}_{\mathcal{M}}^k(n) \geq 2 \cdot (2k+1) \cdot (|A|^k \cdot |\mathcal{M}|^{|\mathcal{M}|} + 2) + n > n \quad (4.6)$$

$$\mathcal{D}_{\mathcal{M}}^k(n+1) \geq 2 \cdot \mathcal{D}_{\mathcal{M}}^k(n) + (2k+1) \cdot (|A|^k \cdot |\mathcal{M}|^{|\mathcal{M}|} + 2) \quad (4.7)$$

Now we are going to prove a proposition showing that every sufficiently large word  $w$  has a proper predecessor  $\hat{w}$  w.r.t.  $\preceq_{\mathcal{M}}^{0,k}$  such that both words are equivalent w.r.t. to  $\mathcal{M}$ .

**Proposition 4.23.** Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $w \in A^*$  with  $|w| \geq (2k+1) \cdot (|A|^k \cdot |\mathcal{M}|^{|\mathcal{M}|} + 2)$ . Then there exists a word  $\hat{w} \in A^*$  such that  $3k+1 \leq |\hat{w}| < |w|$ ,  $\hat{w} \preceq_{\mathcal{M}}^{0,k} w$  and  $\delta^{\hat{w}} = \delta^w$ .

*Proof.* From the length of  $w$  it follows that it can be written as  $w = w'_0 v_1 w'_1 v_2 w'_2 \cdots v_{m'} w'_{m'}$  for  $m' \stackrel{\text{def}}{=} |A|^k \cdot |\mathcal{M}|^{|\mathcal{M}|} + 1$  and words  $v_i \in A^k$ ,  $w'_i \in A^{*\geq k+1}$ . Since  $|A^k| = |A|^k$  there exists some  $v \in A$  such that  $v = v_i$  for at least  $m \stackrel{\text{def}}{=} |\mathcal{M}|^{|\mathcal{M}|} + 1$  words  $v_i$  with  $1 \leq i \leq m'$ . So we can choose suitable words  $w_0, \dots, w_m \in A^{*\geq k+1}$  such that  $w = w_0 v w_1 v w_2 \cdots v w_m$ . For  $0 \leq i \leq m-1$  let  $\delta_i \stackrel{\text{def}}{=} \delta^{w_0 v w_1 \cdots v w_i}$ . Since there are at most  $|\mathcal{M}|^{|\mathcal{M}|}$  different mappings  $\delta^x$  for  $x \in A^*$  there exist indices  $0 \leq i < j \leq m-1$  such that  $\delta_i = \delta_j$ . Hence, for  $\hat{w} \stackrel{\text{def}}{=} w_0 v w_1 \cdots v w_i \cdot v w_{j+1} \cdots v w_m$  (i.e., the word  $w$  after deleting the part  $v w_{i+1} \cdots v w_j$ ) we obtain  $\delta^{\hat{w}} = \delta^w$ . Moreover, with  $x \stackrel{\text{def}}{=} w_0 v w_1 \cdots v w_i$ ,  $z \stackrel{\text{def}}{=} w_{j+1} v w_{j+2} \cdots v w_m$  and  $u \stackrel{\text{def}}{=} w_{i+1} v w_{i+2} \cdots v w_j$  we have  $x, z, u \in A^{*\geq k+1}$ ,  $\hat{w} = xvz$  and  $w = xvuvz$ . This shows  $\hat{w} \prec_{\mathcal{M}}^{0,k} w$  and  $3k+1 \leq |\hat{w}| < |w|$ .  $\square$

**Lemma 4.24.** Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $w, w' \in A^*$  with  $w \preceq_{\mathcal{M}}^{0,k} w'$ . If  $|w'| > \mathcal{D}_{\mathcal{M}}^k(|w|)$  then there exists a word  $\hat{w} \in A^+$  such that  $|\hat{w}| < |w'|$ ,  $w \preceq_{\mathcal{M}}^{0,k} \hat{w} \preceq_{\mathcal{M}}^{0,k} w'$  and  $\delta^{\hat{w}} = \delta^{w'}$ .

*Proof.* We prove the lemma by induction on  $|w'| \geq 0$ . If  $|w'| = 0$  then  $|w'| \leq \mathcal{D}_{\mathcal{M}}^k(|w|)$  and we are done.

As induction hypothesis we assume that there is some  $r \geq 0$  such that the lemma has been shown for all  $w, w' \in A^*$  with  $|w'| \leq r$ . Now let  $w, w' \in A^*$  with  $w \preceq_{\mathcal{M}}^{0,k} w'$  and  $|w'| = r + 1$ . If  $|w'| \leq \mathcal{D}_{\mathcal{M}}^k(|w|)$  then we are done. Otherwise, from  $|w'| > \mathcal{D}_{\mathcal{M}}^k(|w|)$  and (4.6) it follows that  $w \neq w'$ , and therefore  $|w| \geq k$ . We distinguish the following cases.

**Case 1:** Assume that  $w' \in (\mathfrak{p}_k(w)A^{*\geq k+1} \cup \{\varepsilon\}) \cdot w \cdot (A^{*\geq k+1}\mathfrak{s}_k(w) \cup \{\varepsilon\})$ , i.e., there exist words  $x \in (\mathfrak{p}_k(w)A^{*\geq k+1} \cup \{\varepsilon\})$  and  $z \in (A^{*\geq k+1}\mathfrak{s}_k(w) \cup \{\varepsilon\})$  with  $w' = xwz$ . Without loss of generality we assume that  $|x| \geq |z|$ , the other case is treated analogously. From the length of  $w'$  and (4.6) it follows that  $|x| \geq (2k + 1) \cdot (|A|^k \cdot |\mathcal{M}|^{|\mathcal{M}|} + 2)$ . Hence  $x \neq \varepsilon$  and with  $v =_{\text{def}} \mathfrak{p}_k(w)$  it holds that  $x = vu$ ,  $w = v\tilde{w}$  and  $w' = xv\tilde{w}z$  for suitable words  $u \in A^{*\geq k+1}$  and  $\tilde{w} \in A^*$ . So  $|vuv| \geq |x| \geq (2k + 1) \cdot (|A|^k \cdot |\mathcal{M}|^{|\mathcal{M}|} + 2)$  and we can apply Proposition 4.23 to  $vuv$ . We obtain a word  $\hat{x} \in A^*$  such that  $3k + 1 \leq |\hat{x}| < |vuv|$ ,  $\hat{x} \preceq_{\mathcal{M}}^{0,k} vuv$  and  $\delta^{\hat{x}} = \delta^{vuv} = \delta^{xv}$ . By Proposition 4.4,  $\mathfrak{p}_k(\hat{x}) = \mathfrak{p}_k(vuv) = v$  and  $\mathfrak{s}_k(\hat{x}) = \mathfrak{s}_k(vuv) = v$ . Together with  $3k + 1 \leq |\hat{x}|$  this implies  $\hat{x} \in vA^{*\geq k+1}v$ , and therefore  $v \preceq_{\mathcal{M}}^{0,k} \hat{x}$ . Let  $\hat{w} =_{\text{def}} \hat{x}\tilde{w}z$  and note that  $0 < |\hat{w}| < |w'|$  and  $\delta^{\hat{w}} = \delta^{w'}$ . From  $v \preceq_{\mathcal{M}}^{0,k} \hat{x}$  it follows that  $v\tilde{w} \preceq_{\mathcal{M}}^{0,k} \hat{x}\tilde{w}$ . We have already seen that  $\hat{x}$  has the suffix  $v$  which implies that  $\hat{x}\tilde{w}$  has the suffix  $v\tilde{w} = w$ . Therefore,  $\mathfrak{s}_k(\hat{x}\tilde{w}) = \mathfrak{s}_k(w) = v$  and it follows that  $\hat{x}\tilde{w} \preceq_{\mathcal{M}}^{0,k} \hat{x}\tilde{w}z$ . Together with  $\hat{x} \preceq_{\mathcal{M}}^{0,k} vuv$  this yields  $w = v\tilde{w} \preceq_{\mathcal{M}}^{0,k} \hat{x}\tilde{w} \preceq_{\mathcal{M}}^{0,k} \hat{x}\tilde{w}z = \hat{w} \preceq_{\mathcal{M}}^{0,k} vuv\tilde{w}z = xwz = w'$ .

**Case 2:** Assume that  $w' \notin (\mathfrak{p}_k(w)A^{*\geq k+1} \cup \{\varepsilon\}) \cdot w \cdot (A^{*\geq k+1}\mathfrak{s}_k(w) \cup \{\varepsilon\})$ . In particular we have  $w \neq \varepsilon$  since from  $w \neq w'$  it follows that  $w' \in A^{*\geq k+1}$ . Now we can apply Lemma 4.21 to  $w \preceq_{\mathcal{M}}^{0,k} w'$ . We obtain words  $w_1, w_2, w'_1, w'_2 \in A^+$ ,  $v \in W_{\mathcal{M}}^{0,k}$  and  $u \in A^{*\geq k+1}$  with  $w = w_1vw_2$ ,  $w' = w'_1vuw'_2$ ,  $w_1v \preceq_{\mathcal{M}}^{0,k} w'_1v$  and  $vw_2 \preceq_{\mathcal{M}}^{0,k} vw'_2$ .

*Case 2a:* Assume that  $|vuv| \geq (2k + 1) \cdot (|A|^k \cdot |\mathcal{M}|^{|\mathcal{M}|} + 2)$ . Then from Proposition 4.23 we get a word  $\hat{w}' \in A^*$  such that  $3k + 1 \leq |\hat{w}'| < |vuv|$ ,  $\hat{w}' \preceq_{\mathcal{M}}^{0,k} vuv$  and  $\delta^{\hat{w}'} = \delta^{vuv}$ . Let  $\hat{w} =_{\text{def}} w'_1\hat{w}'w'_2$  and observe that  $0 < |\hat{w}| < |w'|$  and  $\delta^{\hat{w}} = \delta^{w'}$ . By Proposition 4.4, from  $w_1v \preceq_{\mathcal{M}}^{0,k} w'_1v$  and  $vw_2 \preceq_{\mathcal{M}}^{0,k} vw'_2$  we obtain  $w_1vw_2 \preceq_{\mathcal{M}}^{0,k} w'_1vw_2$  and  $w'_1vw_2 \preceq_{\mathcal{M}}^{0,k} w'_1vw'_2$ . Therefore, it holds that  $w = w_1vw_2 \preceq_{\mathcal{M}}^{0,k} w'_1vw'_2$ . From  $3k + 1 \leq |\hat{w}'|$  and  $\hat{w}' \preceq_{\mathcal{M}}^{0,k} vuv$  it follows that  $\hat{w}' \in vA^{*\geq k+1}v$  (by Proposition 4.4). This yields  $v \preceq_{\mathcal{M}}^{0,k} \hat{w}'$  and therefore  $w \preceq_{\mathcal{M}}^{0,k} w'_1vw'_2 \preceq_{\mathcal{M}}^{0,k} w'_1\hat{w}'w'_2 = \hat{w} \preceq_{\mathcal{M}}^{0,k} w'_1vuw'_2 = w'$ .

*Case 2b:* Assume that  $|vuv| < (2k + 1) \cdot (|A|^k \cdot |\mathcal{M}|^{|\mathcal{M}|} + 2)$ . Since  $w \neq \varepsilon$  and since  $|w'| > \mathcal{D}_{\mathcal{M}}^k(|w|)$  we obtain from (4.7) that

$$|w'_1| + |w'_2| = |w'| - |vuv| > \mathcal{D}_{\mathcal{M}}^k(|w|) - |vuv| > 2 \cdot \mathcal{D}_{\mathcal{M}}^k(|w| - 1).$$

It follows that at least one of the words  $w'_1, w'_2$  is of length  $> \mathcal{D}_{\mathcal{M}}^k(|w| - 1)$ . Without loss of generality we assume that this holds for  $w'_1$ , i.e.,  $|w'_1v| \geq |w'_1| > \mathcal{D}_{\mathcal{M}}^k(|w| - 1) \geq \mathcal{D}_{\mathcal{M}}^k(|w_1v|)$ . Since  $|w'_1v| < |w'| = r + 1$  we can apply the induction hypothesis to  $w_1v \preceq_{\mathcal{M}}^{0,k} w'_1v$ . We obtain a word  $\hat{w}_1 \in A^+$  with  $|\hat{w}_1| < |w'_1v|$ ,  $w_1v \preceq_{\mathcal{M}}^{0,k} \hat{w}_1 \preceq_{\mathcal{M}}^{0,k} w'_1v$  and  $\delta^{\hat{w}_1} = \delta^{w'_1v}$ . Hence, for  $\hat{w} =_{\text{def}} \hat{w}_1vuw'_2$  it holds that  $0 < |\hat{w}| < |w'|$  and  $\delta^{\hat{w}} = \delta^{w'}$ . From  $w_1v \preceq_{\mathcal{M}}^{0,k} \hat{w}_1$  we obtain that  $\hat{w}_1$  has the suffix  $v$  which in turn implies that  $\hat{w}_1 \preceq_{\mathcal{M}}^{0,k} \hat{w}_1uv$ . This shows  $w_1v \preceq_{\mathcal{M}}^{0,k} \hat{w}_1uv$  and it follows  $w = w_1vw_2 \preceq_{\mathcal{M}}^{0,k} \hat{w}_1uvw_2$ . Together with  $vw_2 \preceq_{\mathcal{M}}^{0,k} vw'_2$  we get  $w \preceq_{\mathcal{M}}^{0,k} \hat{w}_1uvw'_2 = \hat{w}$ . Finally, from  $\hat{w}_1 \preceq_{\mathcal{M}}^{0,k} w'_1v$  it follows that  $\hat{w} = \hat{w}_1uvw'_2 \preceq_{\mathcal{M}}^{0,k} w'_1vuw'_2 = w'$ .  $\square$

**Corollary 4.25.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $w, w' \in A^+$  with  $w \preceq_{\mathcal{M}}^{0,k} w'$ . Then there exists a word  $\hat{w} \in A^+$  such that  $|\hat{w}| \leq \mathcal{D}_{\mathcal{M}}^k(|w|)$ ,  $w \preceq_{\mathcal{M}}^{0,k} \hat{w} \preceq_{\mathcal{M}}^{0,k} w'$  and  $\delta^{\hat{w}} = \delta^{w'}$ .*

*Proof.* This follows when we apply Lemma 4.24 repeatedly.  $\square$



Now we are able to prove that for an arbitrary  $\preceq_{\mathcal{M}}^{0,k}$  chain there exists a  $\preceq_{\mathcal{M}}^{0,k}$  chain of ‘short’ words such that both chains have the same length and their words have the same acceptance behavior w.r.t.  $\mathcal{M}$ . Therefore, when looking for the maximal number of alternations we can restrict ourselves to chains of ‘short’ words. This implies the decidability of the classes  $\mathcal{B}_{1/2}(n)$  and  $\mathcal{L}_{1/2}(n)$  (see Theorem 4.27).

**Theorem 4.26.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $w_0 \preceq_{\mathcal{M}}^{0,k} w_1 \preceq_{\mathcal{M}}^{0,k} \dots \preceq_{\mathcal{M}}^{0,k} w_n$  for  $n \geq 0$  and words  $w_0, \dots, w_n \in A^+$ . Then there exist words  $\hat{w}_0, \dots, \hat{w}_n \in A^+$  such that  $\hat{w}_0 \preceq_{\mathcal{M}}^{0,k} \hat{w}_1 \preceq_{\mathcal{M}}^{0,k} \dots \preceq_{\mathcal{M}}^{0,k} \hat{w}_n$ ,  $|\hat{w}_n| \leq \mathcal{E}_{\mathcal{M}}^k(n)$ ,  $\hat{w}_n \preceq_{\mathcal{M}}^{0,k} w_n$  and  $\delta^{\hat{w}_i} = \delta^{w_i}$  for  $0 \leq i \leq n$ .*

*Proof.* We show the theorem by induction on  $n \geq 0$ . For the induction base let  $n = 0$ . If  $w_0 \leq \mathcal{E}_{\mathcal{M}}^k(0)$  then we are done. Otherwise we have  $|w_0| > \mathcal{D}_{\mathcal{M}}^k(0) \geq (2k+1) \cdot (|A|^k \cdot |\mathcal{M}|^{|\mathcal{M}|} + 2)$ . Now we apply Proposition 4.23 repeatedly to  $w_0$ . We do this as long as the emerging word has a length  $> \mathcal{D}_{\mathcal{M}}^k(0)$ . This procedure yields a word  $\hat{w}_0 \in A^*$  such that  $3k+1 \leq |\hat{w}_0| \leq \mathcal{E}_{\mathcal{M}}^k(0)$ ,  $\hat{w}_0 \preceq_{\mathcal{M}}^{0,k} w_0$  and  $\delta^{\hat{w}_0} = \delta^{w_0}$ . This shows the induction base.

Assume that there is some  $r \geq 0$  such that the theorem has been shown for all  $n \leq r$ . Let  $n =_{\text{def}} r+1$  and  $w_0, \dots, w_{r+1} \in A^+$  with  $w_0 \preceq_{\mathcal{M}}^{0,k} w_1 \preceq_{\mathcal{M}}^{0,k} \dots \preceq_{\mathcal{M}}^{0,k} w_{r+1}$ . By induction hypothesis there exist words  $\hat{w}_0, \dots, \hat{w}_r \in A^+$  such that  $\hat{w}_0 \preceq_{\mathcal{M}}^{0,k} \hat{w}_1 \preceq_{\mathcal{M}}^{0,k} \dots \preceq_{\mathcal{M}}^{0,k} \hat{w}_r$ ,  $|\hat{w}_r| \leq \mathcal{E}_{\mathcal{M}}^k(r)$ ,  $\hat{w}_r \preceq_{\mathcal{M}}^{0,k} w_r$  and  $\delta^{\hat{w}_i} = \delta^{w_i}$  for  $0 \leq i \leq r$ . Hence we have  $\hat{w}_r \preceq_{\mathcal{M}}^{0,k} w_{r+1}$ . From Corollary 4.25 we get a word  $\hat{w}_{r+1} \in A^+$  with  $|\hat{w}_{r+1}| \leq \mathcal{D}_{\mathcal{M}}^k(|\hat{w}_r|) \leq \mathcal{D}_{\mathcal{M}}^k(\mathcal{E}_{\mathcal{M}}^k(r)) = \mathcal{E}_{\mathcal{M}}^k(r+1)$ ,  $\hat{w}_r \preceq_{\mathcal{M}}^{0,k} \hat{w}_{r+1} \preceq_{\mathcal{M}}^{0,k} w_{r+1}$  and  $\delta^{\hat{w}_{r+1}} = \delta^{w_{r+1}}$ .  $\square$

**Theorem 4.27.** *On input of a DFA  $\mathcal{M}$  and  $n \geq 1$  the questions  $L(\mathcal{M}) \in \mathcal{B}_{1/2}(n)$  and  $L(\mathcal{M}) \in \mathcal{L}_{1/2}(n)$  are decidable.*

*Proof.* Let  $k =_{\text{def}} 3 \cdot \mathcal{I}_{\mathcal{M}}$ . By Theorem 4.18, it suffices to find out whether  $m_{\mathcal{M}}^{0,k} < n$  (respectively,  $m_{\mathcal{M}}^{0,0} < n$ ). By definition this means that we have to decide whether there exist words  $w_0, \dots, w_n \in A^+$  with  $w_0 \preceq_{\mathcal{M}}^{0,k} w_1 \preceq_{\mathcal{M}}^{0,k} \dots \preceq_{\mathcal{M}}^{0,k} w_n$  (respectively,  $w_0 \preceq_{\mathcal{M}}^{0,0} w_1 \preceq_{\mathcal{M}}^{0,0} \dots \preceq_{\mathcal{M}}^{0,0} w_n$ ),  $w_0 \in L(\mathcal{M})$  and  $w_{i-1} \in L(\mathcal{M}) \iff w_i \notin L(\mathcal{M})$  for  $1 \leq i \leq n$ . By Theorem 4.26, it suffices to consider chains  $w_0 \preceq_{\mathcal{M}}^{0,k} w_1 \preceq_{\mathcal{M}}^{0,k} \dots \preceq_{\mathcal{M}}^{0,k} w_n$  (respectively,  $w_0 \preceq_{\mathcal{M}}^{0,0} w_1 \preceq_{\mathcal{M}}^{0,0} \dots \preceq_{\mathcal{M}}^{0,0} w_n$ ) for words  $w_i \in A^+$  with  $|w_i| \leq \mathcal{E}_{\mathcal{M}}^k(n)$ . The theorem follows since  $\mathcal{E}_{\mathcal{M}}^k(n)$  is computable and finite, and since the questions  $w_i \in L(\mathcal{M})$ ,  $w_i \preceq_{\mathcal{M}}^{0,k} w_{i+1}$  and  $w_i \preceq_{\mathcal{M}}^{0,0} w_{i+1}$  are decidable.  $\square$

**Theorem 4.28.** *The following functions are computable on input of a DFA  $\mathcal{M}$ .*

$$\begin{aligned} m_{\mathcal{B}}(\mathcal{M}) &=_{\text{def}} \inf \{ n \geq 1 \mid L(\mathcal{M}) \in \mathcal{B}_{1/2}(n) \} \\ m_{\mathcal{L}}(\mathcal{M}) &=_{\text{def}} \inf \{ n \geq 1 \mid L(\mathcal{M}) \in \mathcal{L}_{1/2}(n) \} \end{aligned}$$

*Proof.* For this proof we need the facts that  $\mathcal{B}_1$  and  $\mathcal{L}_1$  are decidable. For  $\mathcal{B}_1$  this was first shown in [Kna83] with an algebraic approach (see also [Ste85b, CH91]). The decidability of  $\mathcal{L}_1$  is due to [Sim75].

Now the computation of  $m_{\mathcal{B}}(\mathcal{M})$  is as follows. First of all we test whether  $L(\mathcal{M})$  belongs to  $\mathcal{B}_1$ . If  $L(\mathcal{M}) \notin \mathcal{B}_1$  then  $m_{\mathcal{B}}(\mathcal{M}) = \infty$  and we are done. Otherwise it follows that the value of  $m_{\mathcal{B}}(\mathcal{M})$  is finite. In this case we carry out tests  $L(\mathcal{M}) \in \mathcal{B}_{1/2}(n)$  for  $n = 1, 2, \dots$  (for

these tests we use the procedure described in the proof of Theorem 4.27). The first  $n$  that passes the test is the value of  $m_{\mathcal{B}}(\mathcal{M})$ . Since  $m_{\mathcal{B}}(\mathcal{M})$  is finite such an  $n$  actually exists, and therefore our algorithm terminates. Analogously we show that  $m_{\mathcal{L}}(\mathcal{M})$  is computable.  $\square$

From the proof of Theorem 4.28 we cannot derive a bound for the number of steps needed to compute the functions  $m_{\mathcal{B}}$  and  $m_{\mathcal{L}}$ . However, such a bound can be easily established with the following proposition at hand.

**Proposition 4.29.** *For every minimal DFA  $\mathcal{M}$  and  $k =_{\text{def}} 3 \cdot \mathcal{I}_{\mathcal{M}}$  the following holds.*

$$\begin{aligned} m_{\mathcal{B}}(\mathcal{M}) \geq 2^{2|A|^{k+2}(k+1)^2|\mathcal{M}|} + 3 &\implies m_{\mathcal{B}}(\mathcal{M}) = \infty \\ m_{\mathcal{L}}(\mathcal{M}) \geq 2^{|A|^2 \cdot |\mathcal{M}|} + 3 &\implies m_{\mathcal{L}}(\mathcal{M}) = \infty \end{aligned}$$

*Proof.* Recall the  $k$ -embedding  $\triangleleft_k$  from [Ste85a] (see also [Sch01, section 2.7] and [Sch01, Definition 2.1] for a discussion and for an equivalent definition). It is easy to see that if  $w \triangleleft_v^{0,k} w'$  for  $w, w' \in A^{*\geq k+1}$  and  $v \in A^k$  then  $w \triangleleft_k w'$ . Since  $\triangleleft_k$  is reflexive and transitive we have  $w \triangleleft_k w'$  for all  $w, w' \in A^{*\geq k+1}$  with  $w \triangleleft_{\mathcal{M}}^{0,k} w'$ . Analogously we see that  $w \preceq w'$  for all  $w, w' \in A^+$  with  $w \triangleleft_{\mathcal{M}}^{0,0} w'$ .

Assume that  $m_{\mathcal{B}}(\mathcal{M}) \geq 2^{2|A|^{k+2}(k+1)^2|\mathcal{M}|} + 3$  and  $m_{\mathcal{B}}(\mathcal{M}) < \infty$ . First of all, by the definition of  $m_{\mathcal{B}}(\mathcal{M})$  this implies  $L(\mathcal{M}) \in \mathcal{B}_{1/2}(m_{\mathcal{B}}(\mathcal{M})) \subseteq \mathcal{B}_1$ , i.e.,  $L(\mathcal{M})$  belongs to dot-depth one. From [Ste85a, Proposition 4.1] we obtain that  $L(\mathcal{M})$  belongs to some level  $\leq |\mathcal{M}|^3 \leq k$  of dot-depth one (see [Ste85a, section 3.3] for a definition of these levels).

Our choice of  $k$  and Theorem 4.18 imply  $m_{\mathcal{M}}^{0,k} = m_{\mathcal{B}}(\mathcal{M}) - 1 \geq 2^{2|A|^{k+2}(k+1)^2|\mathcal{M}|} + 2$ . This means that for  $n =_{\text{def}} m_{\mathcal{B}}(\mathcal{M}) - 1$  there exists a chain  $w_0 \triangleleft_{\mathcal{M}}^{0,k} w_1 \triangleleft_{\mathcal{M}}^{0,k} \dots \triangleleft_{\mathcal{M}}^{0,k} w_n$  with  $w_i \in L(\mathcal{M}) \iff i \equiv 0 \pmod{2}$ . In particular  $w_0 \neq w_1$ , and therefore  $w_1, \dots, w_n \in A^{*\geq k+1}$ . It follows that  $w_1 \triangleleft_k w_2 \triangleleft_k \dots \triangleleft_k w_n$  which is a  $\triangleleft_k$  chain with  $n - 1 > 2^{2|A|^{k+2}(k+1)^2|\mathcal{M}|}$  alternations with respect to  $L(\mathcal{M})$  (in [Ste85a] such chains are called  $k$ -towers). From [Ste85a, Theorem 3.3] we obtain that  $L(\mathcal{M})$  does not belong to some level  $\leq k$  of dot-depth one. This is a contradiction which proves the first fact of the proposition.

For the second fact we assume  $m_{\mathcal{L}}(\mathcal{M}) \geq 2^{|A|^2 \cdot |\mathcal{M}|} + 3$  and  $m_{\mathcal{L}}(\mathcal{M}) < \infty$ . By the definition of  $m_{\mathcal{L}}(\mathcal{M})$  this implies  $L(\mathcal{M}) \in \mathcal{L}_1$  which is equivalent to saying that  $L(\mathcal{M})$  is a piecewise testable language.

By Theorem 4.18 we have  $m_{\mathcal{M}}^{0,0} = m_{\mathcal{L}}(\mathcal{M}) - 1 = 2^{|A|^2 \cdot |\mathcal{M}|} + 2$ . So for  $n =_{\text{def}} m_{\mathcal{L}}(\mathcal{M}) - 1$  there exists a chain  $w_0 \triangleleft_{\mathcal{M}}^{0,0} w_1 \triangleleft_{\mathcal{M}}^{0,0} \dots \triangleleft_{\mathcal{M}}^{0,0} w_n$  with  $w_i \in L(\mathcal{M}) \iff i \equiv 0 \pmod{2}$ . In particular  $w_0 \neq w_1$ , and therefore  $w_1, \dots, w_n \in A^+$ . It follows that  $w_1 \preceq w_2 \preceq \dots \preceq w_n$  which is a  $\preceq$  chain with  $n - 1 > 2^{|A|^2 \cdot |\mathcal{M}|}$  alternations with respect to  $L(\mathcal{M})$  (in [Ste85a] such chains are called towers). From [Ste85a, Theorem 2.1] we obtain that  $L(\mathcal{M})$  is not piecewise testable. This is a contradiction.  $\square$

### 4.3 The Levels 3/2

We prove effective forbidden-pattern characterizations for the levels 3/2 of the DDH and STH. For the STH this was first shown in [PW97], the result for the DDH is due to

[GS00a]. However, in this section we prove both results with a new technique which uses word extensions. We proceed analogously to section 4.1 where we considered the levels 1/2.

First of all we define the word extensions  $\preceq_{\mathcal{M}}^{1,k}$  which can be also considered as binary, reflexive, transitive and antisymmetric relations on the set of words. In subsection 4.3.2 we show that the  $\preceq_{\mathcal{M}}^{1,k}$  upward closure of a nonempty word (i.e., the  $\preceq_{\mathcal{M}}^{1,k}$  co-ideal generated by a nonempty word) is in  $\mathcal{B}_{3/2}$ . Then in subsection 4.3.3 we prove: If the language accepted by some DFA  $\mathcal{M}$  is a  $\preceq_{\mathcal{M}}^{1,k}$  co-ideal then this co-ideal is finitely generated. Together with the result of subsection 4.3.2 this implies that these regular  $\preceq_{\mathcal{M}}^{1,k}$  co-ideals are in  $\mathcal{B}_{3/2}$ .

Note that this differs from the procedure for the levels 1/2 in section 4.1. There we showed more, namely that the set of words together with  $\preceq_{\mathcal{M}}^{0,k}$  is a well partial ordered set. Unfortunately, this does not hold for  $\preceq_{\mathcal{M}}^{1,k}$  (see the remark after Theorem 4.46). Therefore, in order to prove that  $\preceq_{\mathcal{M}}^{1,k}$  co-ideals are finitely generated we have to restrict ourselves to the regular case.

Finally, in this section we show that the languages  $L(\mathcal{M})$  that belong to the forbidden-pattern class  $\mathcal{FP}(\mathbb{P}_1^{\mathcal{B}})$  are regular  $\preceq_{\mathcal{M}}^{1,k}$  co-ideals, and therefore we obtain  $\mathcal{FP}(\mathbb{P}_1^{\mathcal{B}}) \subseteq \mathcal{B}_{3/2}$ . The reverse inclusion is known from the pattern theory in chapter 3, and therefore we obtain the effective forbidden-pattern characterization  $\mathcal{FP}(\mathbb{P}_1^{\mathcal{B}}) = \mathcal{B}_{3/2}$ . In particular this shows that the membership problem for  $\mathcal{B}_{3/2}$  is decidable in nondeterministic logarithmic space. Again, in this section we treat the DDH and the STH in parallel.

### 4.3.1 Definition of $\preceq_{\mathcal{M}}^{1,k}$ Word Extensions

We proceed analogously to subsection 4.1.1 where the levels 1/2 of the DDH and STH were considered. Here we introduce the notion of  $\preceq_{\mathcal{M}}^{1,k}$  extensions. They are defined in such a way that  $x \preceq_{\mathcal{M}}^{1,k} y$  means that the word  $y$  results from the word  $x$  by a sequence of extensions  $\prec_{v_1}^{1,k}, \prec_{v_2}^{1,k}, \dots, \prec_{v_m}^{1,k}$ . It is important that we do not fix the context word but we allow several words  $v_0, v_1, \dots, v_m$ . Since these words have to satisfy certain conditions we define the set of possible context words for  $\preceq_{\mathcal{M}}^{1,k}$  extensions in Definition 4.31 below.

We start with the definition of two bounding functions which will be needed in the proofs below.

**Definition 4.30.** *Let  $\mathcal{M}$  be a DFA and  $k \geq 0$ .*

$$\begin{aligned} \mathcal{F}_{\mathcal{M}}^k(m) &=_{\text{def}} \begin{cases} k+1 & : \text{ if } m = 0 \\ \mathcal{F}_{\mathcal{M}}^k(m-1) \cdot (\mathcal{I}_{\mathcal{M}} + 1) \cdot (|\mathcal{M}| \cdot |A|^{\mathcal{I}_{\mathcal{M}} \cdot \mathcal{F}_{\mathcal{M}}^k(m-1)+1}) & : \text{ if } m > 0 \end{cases} \\ \mathcal{C}_{\mathcal{M}}^k &=_{\text{def}} \mathcal{F}_{\mathcal{M}}^k(|A|^{k+1}) \end{aligned}$$

It can be easily verified that  $\mathcal{F}_{\mathcal{M}}^k(\cdot)$  is a positive and monotone increasing function.

**Definition 4.31.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $n =_{\text{def}} \mathcal{I}_{\mathcal{M}}$ . Then we define the following set of context words.*

$$\mathcal{W}_{\mathcal{M}}^{1,k} =_{\text{def}} \left\{ v \in A^{*\leq \mathcal{C}_{\mathcal{M}}^k} \mid \begin{array}{l} v = r_1 r_2 \cdots r_n = l_n l_{n-1} \cdots l_1 \text{ for suitable } l_i, r_i \in A^{*\geq k+1} \text{ with } \\ \alpha_k(A^{-1}l_i), \alpha_k(r_i A^{-1}) \subsetneq \alpha_k(v) \text{ and } \alpha_k(r_i) = \alpha_k(l_i) = \alpha_k(v) \end{array} \right\}$$

We want to see that the set of context words  $W_{\mathcal{M}}^{1,k}$  is nonempty for every DFA  $\mathcal{M}$  and every  $k \geq 0$ . For this let  $n =_{\text{def}} \mathcal{I}_{\mathcal{M}}$  and let  $a$  be an arbitrary letter from  $A$ . From  $n \geq 1$ ,  $|\mathcal{M}| \geq 1$  and  $|A| \geq 2$  it follows that  $\mathcal{C}_{\mathcal{M}}^k \geq n(k+1)$ . Therefore, the word  $v =_{\text{def}} a^{n(k+1)}$  is an element of  $A^{*\leq \mathcal{C}_{\mathcal{M}}^k}$ . Moreover, if we define  $r_i = l_i = a^{k+1}$  then  $v$  can be written as  $v = r_1 \cdots r_n = l_1 \cdots l_n$ . Finally, from  $\alpha_k(r_i) = \alpha_k(l_i) = \{a^{k+1}\} = \alpha_k(v)$  and  $\alpha_k(A^{-1}l_i) = \alpha_k(r_i A^{-1}) = \emptyset$  it follows that  $v$  is an element of  $W_{\mathcal{M}}^{1,k}$ .

**Definition 4.32.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $y, y' \in A^*$ .*

$$y \preceq_{\mathcal{M}}^{1,k} y' \iff_{\text{def}} \text{there exist an } m \geq 0, \text{ words } x_0, \dots, x_m \in A^* \text{ and } v_1, \dots, v_m \in W_{\mathcal{M}}^{1,k} \\ \text{such that } y = x_0 \prec_{v_1}^{1,k} x_1 \prec_{v_2}^{1,k} \cdots \prec_{v_m}^{1,k} x_m = y'$$

The definition of the set of context words  $W_{\mathcal{M}}^{1,k}$  seems a bit arbitrary. However, it is guided by the following ideas (which are not obvious at the moment but which will be proved in this section).

1. each  $v \in W_{\mathcal{M}}^{1,k}$  contains a factor  $u$  such that  $u$  is an idempotent for  $\mathcal{M}$  and  $\alpha_k(u) = \alpha_k(v)$
2. words from  $W_{\mathcal{M}}^{1,k}$  are short, i.e., they are of bounded length
3. no  $v \in W_{\mathcal{M}}^{1,k}$  contains a proper factor  $v' \in W_{\mathcal{M}}^{1,k}$  unless  $\alpha_k(v') \subsetneq \alpha_k(v)$
4.  $W_{\mathcal{M}}^{1,k}$  is a  $\preceq_{\mathcal{M}}^{1,k}$  order ideal, i.e., if  $v \in W_{\mathcal{M}}^{1,k}$  and  $v' \preceq_{\mathcal{M}}^{1,k} v$  then  $v' \in W_{\mathcal{M}}^{1,k}$

The first property is used to prove that languages  $L$  from the forbidden-pattern class  $\mathcal{FP}(\mathbb{P}_1^{\mathbb{B}})$  are regular  $\preceq_{\mathcal{M}}^{1,k}$  co-ideals where  $\mathcal{M}$  is a DFA with  $L = L(\mathcal{M})$ . The second one is needed to show that these co-ideals are finitely generated. The remaining properties allow to change the order of context words  $v_i$  in sequences of  $\prec_v^{1,k}$  extensions. This helps to show that the  $\preceq_{\mathcal{M}}^{1,k}$  upward closure of a nonempty word is in  $\mathcal{B}_{3/2}$ .

In the following proposition we state some basic results about  $\preceq_{\mathcal{M}}^{1,k}$  extensions. In particular it holds that they are stable word extensions which preserve the  $k$ -prefix and the  $k$ -suffix.

**Proposition 4.33.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $v, w, w' \in A^*$ . Then the following holds.*

1. If  $w \preceq_{\mathcal{M}}^{1,k} w'$  then  $\mathfrak{p}_k(w) = \mathfrak{p}_k(w')$  and  $\mathfrak{s}_k(w) = \mathfrak{s}_k(w')$ .
2. If  $w \preceq_{\mathcal{M}}^{1,k} w'$  then  $xwz \preceq_{\mathcal{M}}^{1,k} xw'z$  for all  $x, z \in A^*$ .
3. If  $w \preceq_{\mathcal{M}}^{1,k} w'$  then  $\alpha_k(w) = \alpha_k(w')$ .
4. If  $v \preceq_{\mathcal{M}}^{1,k} w$  and  $w \in vA^{*\geq k+1}$  then  $v \prec_v^{1,k} w$ .

*Proof.* Trivially, if  $w = w'$  then the statements 1, 2 and 3 hold. If  $w \neq w'$  then by definition there exist an  $m \geq 1$ , words  $x_0, \dots, x_m \in A^*$  and context words  $v_1, \dots, v_m \in W_{\mathcal{M}}^{1,k}$  such that  $w = x_0 \prec_{v_1}^{1,k} x_1 \prec_{v_2}^{1,k} \cdots \prec_{v_m}^{1,k} x_m = w'$ . Since elements of  $W_{\mathcal{M}}^{1,k}$  are of length  $\geq k+1$  we have  $|v_i| \geq k+1$  for  $1 \leq i \leq m$ . Therefore, the statements 1, 2 and 3 follow from Proposition 1.19.

If  $v \preceq_{\mathcal{M}}^{1,k} w$  and  $w \in vA^{*\geq k+1}$  then there exists a  $w' \in A^{*\geq k+1}$  with  $w = vw'v$ . From statement 3 it follows that  $\alpha_k(v) = \alpha_k(w) = \alpha_k(vw'v)$ . This shows  $v \prec_v^{1,k} w$ .  $\square$

One of the nice properties of  $W_{\mathcal{M}}^{1,k}$  is the following: No  $v \in W_{\mathcal{M}}^{1,k}$  contains a proper factor  $v' \in W_{\mathcal{M}}^{1,k}$  unless  $\alpha_k(v') \subsetneq \alpha_k(v)$ . We will use this property in subsection 4.3.2 to change the order of context words  $v_i$  in sequences of  $\prec_v^{1,k}$  extensions.

**Proposition 4.34.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $v, v' \in W_{\mathcal{M}}^{1,k}$  such that  $v'$  is a factor of  $v$  and  $\alpha_k(v) = \alpha_k(v')$ . Then it holds that  $v = v'$ .*

*Proof.* Let  $n =_{\text{def}} \mathcal{I}_{\mathcal{M}}$ . By the definition of  $W_{\mathcal{M}}^{1,k}$  we have

$$\begin{aligned} v &= r_1 r_2 \cdots r_n = l_n l_{n-1} \cdots l_1 \quad \text{and} \\ v' &= r'_1 r'_2 \cdots r'_n = l'_n l'_{n-1} \cdots l'_1 \end{aligned}$$

for suitable words  $l_i, r_i, l'_i, r'_i \in A^{*\geq k+1}$  that have the properties stated in Definition 4.31. Since  $v'$  is a factor of  $v$  there exist  $x, z \in A^*$  such that  $v = xv'z$ . Now we will compare the decompositions  $v = r_1 r_2 \cdots r_n$  and  $v = xr'_1 r'_2 \cdots r'_n z$ .

We obtain that  $|xr'_1| \geq |r_1|$  since otherwise  $r'_1$  would be a factor of  $r_1 A^{-1}$  and we would obtain  $\alpha_k(r'_1) \subsetneq \alpha_k(v)$ . This implies  $|xr'_1 r'_2| \geq |r_1 r_2|$  since otherwise  $r'_2$  would be a factor of  $r_2 A^{-1}$  and we would obtain  $\alpha_k(r'_2) \subsetneq \alpha_k(v)$ . Analogously we obtain  $|xr'_1 \cdots r'_i| \geq |r_1 \cdots r_i|$  for  $1 \leq i \leq n$ . This means  $|xr'_1 \cdots r'_n| \geq |v|$  and it follows that  $z = \varepsilon$ .

Analogously we show  $x = \varepsilon$  (here we argue with the decompositions  $v = l_n l_{n-1} \cdots l_1$  and  $v = x l'_n l'_{n-1} \cdots l'_1 z$ ). This shows  $v = v'$ .  $\square$

With the following proposition we show in particular that  $W_{\mathcal{M}}^{1,k}$  is a  $\preceq_{\mathcal{M}}^{1,k}$  order ideal, i.e., if  $v \in W_{\mathcal{M}}^{1,k}$  and  $v' \preceq_{\mathcal{M}}^{1,k} v$  then  $v' \in W_{\mathcal{M}}^{1,k}$ . Actually we show more than this since we assume a condition weaker than  $v' \preceq_{\mathcal{M}}^{1,k} v$ .

**Proposition 4.35.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $v \in W_{\mathcal{M}}^{1,k}$ . If there exist words  $x, w', z \in A^*$  and  $v' \in W_{\mathcal{M}}^{1,k}$  with  $v = xv'w'v'z$  and  $\alpha_k(v'w'v') \subseteq \alpha_k(v') \subsetneq \alpha_k(v)$  then  $xv'z \in W_{\mathcal{M}}^{1,k}$  and  $\alpha_k(xv'z) = \alpha_k(v)$ .*

*Proof.* Let  $n =_{\text{def}} \mathcal{I}_{\mathcal{M}}$  and note that  $n \geq 16$ . By the definition of  $W_{\mathcal{M}}^{1,k}$  we have

$$\begin{aligned} v &= r_1 r_2 \cdots r_n = l_n l_{n-1} \cdots l_1 \\ v' &= r'_1 r'_2 \cdots r'_n = l'_n l'_{n-1} \cdots l'_1 \end{aligned}$$

for suitable words  $l_i, r_i, l'_i, r'_i \in A^{*\geq k+1}$  that have the properties stated in Definition 4.31. We want to compare the following decompositions of  $v$ .

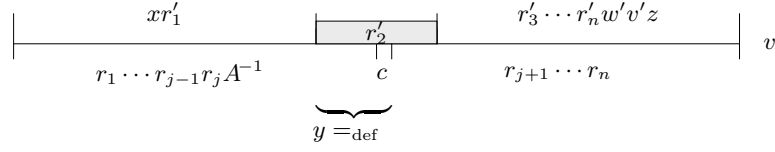
$$v = r_1 r_2 \cdots r_n \tag{4.8}$$

$$= xr'_1 r'_2 r'_3 \cdots r'_n w' v' z \tag{4.9}$$

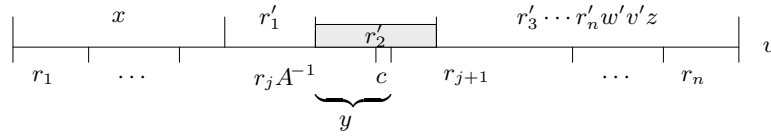
More precisely, we want to investigate the position of the factor  $r'_2$  in the decomposition (4.8). For this we prove the following claim.

**Claim.** *There exists some  $j$  such that the factor  $r'_2$  of the decomposition (4.9) appears in the factor  $r_j A^{-1}$  of the decomposition (4.8).*

Assume that the claim does not hold. Then there exists some  $j$  with  $1 \leq j \leq n$  such that  $r'_2$  overlaps the last letter of the factor  $r_j$ , i.e., the letter  $c =_{\text{def}} \mathfrak{s}_1(r_j)$ . So we have the following situation.

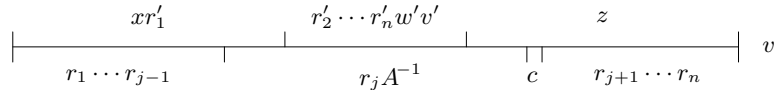


Define  $y$  as shown in the picture above, and note that  $y \in A^+$ . Observe that  $r'_1$  cannot start earlier than  $r_j$ , since otherwise  $r_j$  would be a factor of  $r'_1 r'_2$  and it would follow that  $\alpha_k(r_j) \subseteq \alpha_k(r'_1 r'_2) = \alpha_k(v') \subsetneq \alpha_k(v)$ . So we obtain that  $r'_1 y$  is a factor of  $r_j$ .



On the one hand it holds that  $\mathfrak{s}_{k+1}(r_j) \notin \alpha_k(r_j A^{-1})$  by Definition 4.31. On the other hand  $\mathfrak{s}_{k+1}(r_j)$  is certainly an element of  $\alpha_k(r'_1 r'_2) = \alpha_k(v')$ . Since  $|y| \geq 1$ ,  $r'_1$  is a factor of  $r_j A^{-1}$  which implies  $\alpha_k(v') = \alpha_k(r'_1) \subseteq \alpha_k(r_j A^{-1})$ . It follows that  $\mathfrak{s}_{k+1}(r_j) \in \alpha_k(r_j A^{-1})$  which is a contradiction. This shows our claim.

As an easy consequence of the claim we obtain that even the factor  $r'_2 r'_3 \cdots r'_n w'v'$  of the decomposition (4.9) appears in the factor  $r_j A^{-1}$  of the decomposition (4.8). Otherwise  $\mathfrak{s}_{k+1}(r_j)$  would be a factor of  $r'_2 r'_3 \cdots r'_n w'v'$  and it would follow that  $\mathfrak{s}_{k+1}(r_j) \in \alpha_k(v') \subseteq \alpha_k(r_j A^{-1})$  which is a contradiction. Hence we have reached the following situation.



Let  $\bar{r}_j$  be the word that is obtained from  $r_j$  if one deletes the factor  $w'v'$ . Thus we obtain  $xv'z = r_1 \cdots r_{j-1} \bar{r}_j r_{j+1} \cdots r_n$ . Since  $\mathfrak{s}_k(r'_n) = \mathfrak{s}_k(v')$  and since  $w'v'$  is a factor of  $r_j A^{-1}$  we obtain

$$\mathfrak{p}_{k+1}(\bar{r}_j) = \mathfrak{p}_{k+1}(r_j) \quad \text{and} \quad \mathfrak{s}_{k+1}(\bar{r}_j) = \mathfrak{s}_{k+1}(r_j). \quad (4.10)$$

Together with  $\alpha_k(v'w'v') = \alpha_k(v') = \alpha_k(r'_2)$  it follows that

$$\alpha_k(\bar{r}_j A^{-1}) = \alpha_k(r_j A^{-1}) \quad \text{and} \quad \alpha_k(\bar{r}_j) = \alpha_k(r_j). \quad (4.11)$$

From (4.10) and (4.11) we get  $\alpha_k(xv'z) = \alpha_k(r_1 \cdots r_{j-1} \bar{r}_j r_{j+1} \cdots r_n) = \alpha_k(r_1 \cdots r_n) = \alpha_k(v)$  and  $\mathfrak{s}_{k+1}(\bar{r}_j) \notin \alpha_k(\bar{r}_j A^{-1})$ . Now let  $\bar{r}_i =_{\text{def}} r_i$  for  $1 \leq i \leq n$  with  $i \neq j$ . We obtain  $xv'z = \bar{r}_1 \bar{r}_2 \cdots \bar{r}_n$ ,  $\bar{r}_i \in A^{\geq k+1}$ ,  $\alpha_k(\bar{r}_i) = \alpha_k(xv'z)$  and  $\mathfrak{s}_{k+1}(\bar{r}_i) \notin \alpha_k(\bar{r}_i A^{-1})$  for  $1 \leq i \leq n$ .

Analogously we show that there are words  $\bar{l}_i \in A^{\geq k+1}$  such that  $xv'z = \bar{l}_n \bar{l}_{n-1} \cdots \bar{l}_1$  with  $\alpha_k(\bar{l}_i) = \alpha_k(xv'z)$  and  $\mathfrak{p}_{k+1}(\bar{l}_i) \notin \alpha_k(A^{-1} \bar{l}_i)$  for  $1 \leq i \leq n$ . This shows  $xv'z \in \mathbb{W}_{\mathcal{M}}^{1,k}$ .  $\square$

### 4.3.2 The $\preceq_{\mathcal{M}}^{1,k}$ Upward Closure of a Word

We show that the  $\preceq_{\mathcal{M}}^{1,0}$  upward closure (respectively,  $\preceq_{\mathcal{M}}^{1,k}$  upward closure) of a word is in  $\mathcal{L}_{3/2}$  (respectively,  $\mathcal{B}_{3/2}$  for  $k \geq 0$ ). Most of this subsection is devoted to the preparation for Theorem 4.40 which allows to rearrange context words in sequences of  $\prec_v^{1,k}$  extensions. More precisely, the theorem says that if  $y \preceq_{\mathcal{M}}^{1,k} y'$  then there exists pairwise different context words  $v_1, v_2, \dots, v_n \in W_{\mathcal{M}}^{1,k}$  with  $|\alpha_k(v_1)| \geq |\alpha_k(v_2)| \geq \dots \geq |\alpha_k(v_n)|$  such that there is a sequence of the following form (where suitable words are at the  $*$  positions).

$$y \prec_{v_1}^{1,k} * \prec_{v_1}^{1,k} * \prec_{v_1}^{1,k} \dots \prec_{v_1}^{1,k} * \prec_{v_2}^{1,k} * \prec_{v_2}^{1,k} * \prec_{v_2}^{1,k} \dots \prec_{v_2}^{1,k} \dots \prec_{v_n}^{1,k} * \prec_{v_n}^{1,k} * \prec_{v_n}^{1,k} \dots \prec_{v_n}^{1,k} y' \quad (4.12)$$

Finally, from this theorem we derive that  $\langle y \rangle_{\preceq_{\mathcal{M}}^{1,0}} \in \mathcal{L}_{3/2}$  and  $\langle y \rangle_{\preceq_{\mathcal{M}}^{1,k}} \in \mathcal{B}_{3/2}$  for  $y \in A^+$  and  $k \geq 0$ . We start with a lemma showing that every  $\prec_v^{1,k}$  chain of length two can be transformed into this form (possibly at the cost of a lengthening of the chain).

**Lemma 4.36.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$ ,  $y_1, y_2, y_3 \in A^*$  and  $v_1, v_2 \in W_{\mathcal{M}}^{1,k}$  such that  $|\alpha_k(v_1)| < |\alpha_k(v_2)|$  and  $y_1 \prec_{v_1}^{1,k} y_2 \prec_{v_2}^{1,k} y_3$ . Then at least one of the following statements holds.*

1. *There exists a  $y' \in A^*$  such that  $y_1 \prec_{v_2}^{1,k} y' \prec_{v_1}^{1,k} y_3$ .*
2. *There exist  $y', y'' \in A^*$ ,  $v' \in W_{\mathcal{M}}^{1,k}$  with  $\alpha_k(v') = \alpha_k(v_2)$  and  $y_1 \prec_{v'}^{1,k} y' \prec_{v_1}^{1,k} y'' \prec_{v_1}^{1,k} y_3$ .*

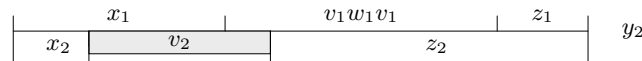
*Proof.* Let  $y_1, y_2, y_3, v_1, v_2$  as in the lemma and let  $n =_{\text{def}} \mathcal{I}_{\mathcal{M}}$ . Then for  $i \in \{1, 2\}$  there exist words  $x_i, z_i \in A^*$  and  $w_i \in A^{*\geq k+1}$  such that  $y_i = x_i v_i z_i$ ,  $y_{i+1} = x_i v_i w_i v_i z_i$  and  $\alpha_k(v_i w_i v_i) \subseteq \alpha_k(v_i)$ . We want to compare the decompositions  $y_2 = x_1 v_1 w_1 v_1 z_1$  and  $y_2 = x_2 v_2 z_2$ . More precisely we want to clarify the position of  $v_2$  in the former decomposition. Observe that  $v_2$  cannot be a factor of  $v_1 w_1 v_1$  since  $\alpha_k(v_1 w_1 v_1) \subseteq \alpha_k(v_1)$  and  $|\alpha_k(v_1)| < |\alpha_k(v_2)|$ .

**Case 1:** Assume that  $v_1 w_1 v_1$  and  $v_2$  does not overlap. Then  $v_2$  is a factor of  $x_1$  or it is a factor of  $z_1$ . Without loss of generality we assume the former.



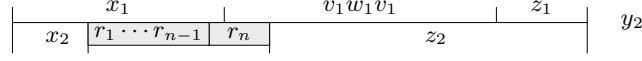
Therefore,  $x_1 = x_2 v_2 x'_1$  for a suitable word  $x'_1 \in A^*$ . It follows that  $y_1 = x_2 v_2 x'_1 v_1 z_1$  and  $y_3 = x_2 v_2 w_2 v_2 x'_1 v_1 w_1 v_1 z_1$ . Now it is easy to see that with  $y' =_{\text{def}} x_2 v_2 w_2 v_2 x'_1 v_1 z_1$  we get  $y_1 \prec_{v_2}^{1,k} y' \prec_{v_1}^{1,k} y_3$ , i.e., statement 1 of the lemma.

**Case 2:** Assume that  $v_1 w_1 v_1$  and  $v_2$  overlap, but none of them is a factor of the other one. Without loss of generality we may assume that  $v_2$  appears earlier than  $v_1 w_1 v_1$ .

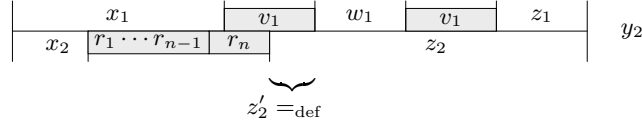


Since  $v_1, v_2 \in W_{\mathcal{M}}^{1,k}$  it holds that  $v_1 = l_n \dots l_1$  and  $v_2 = r_1 \dots r_n$  for suitable words  $l_1, \dots, l_n, r_1, \dots, r_n \in A^{*\geq k+1}$  with  $\alpha_k(A^{-1}l_i) \subsetneq \alpha_k(l_i) = \alpha_k(v_1)$  and  $\alpha_k(r_i A^{-1}) \subsetneq \alpha_k(r_i) = \alpha_k(v_2)$  for  $1 \leq i \leq n$ .

From  $\alpha_k(r_n) = \alpha_k(v_2)$  and  $|\alpha_k(v_1w_1v_1)| = |\alpha_k(v_1)| < |\alpha_k(v_2)|$  it follows that  $r_n$  is not a factor of  $v_1w_1v_1$ . Hence  $r_n$  starts earlier than  $v_1w_1v_1$ .

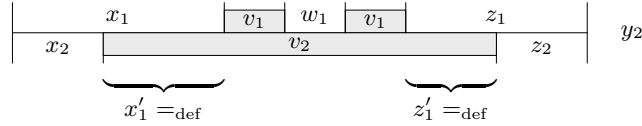


For the prefixes  $x_1v_1$  and  $x_2r_1 \cdots r_n$  of  $y_2$  we want to show that  $|x_1v_1| \geq |x_2r_1 \cdots r_n|$ . If  $|x_1v_1| < |x_2r_1 \cdots r_n|$  then  $v_1$  is a factor of  $r_nA^{-1}$ , and from  $|v_1| \geq k+1$  it follows that  $\mathfrak{s}_{k+1}(r_n) \in \alpha_k(v_1w_1v_1)$ . This implies  $\mathfrak{s}_{k+1}(r_n) \in \alpha_k(v_1w_1v_1) = \alpha_k(v_1) \subseteq \alpha_k(r_nA^{-1})$  which is a contradiction. This shows  $|x_1v_1| \geq |x_2r_1 \cdots r_n|$  and we have reached the following situation.



Let  $z'_2$  as in the picture and observe that  $y_1 = x_2v_2z'_2z_1$  and  $y_3 = x_2v_2w_2v_2z'_2w_1v_1z_1$ . Since  $v_2z'_2$  has  $v_1$  as a suffix we obtain  $y_1 <_{v_2}^{1,k} y' <_{v_1}^{1,k} y_3$  for  $y' =_{\text{def}} x_2v_2w_2v_2z'_2z_1$ . Therefore, statement 1 holds also in case 2 and it remains to consider the following case.

**Case 3:** Assume that  $v_1w_1v_1$  is a factor of  $v_2$  and define  $x'_1, z'_1$  as in the picture below.



From Proposition 4.35 it follows that  $v' =_{\text{def}} x'_1v_1z'_1$  is an element of  $W_{\mathcal{M}}^{1,k}$  with  $\alpha_k(v') = \alpha_k(v_2)$ . Since  $v'$  and  $v_2$  have  $x'_1v_1$  as a prefix and  $v_1z'_1$  as a suffix it holds that  $\mathfrak{p}_k(v') = \mathfrak{p}_k(v_2)$  and  $\mathfrak{s}_k(v') = \mathfrak{s}_k(v_2)$ . Therefore, we get  $\alpha_k(v'w_2v') = \alpha_k(v_2w_2v_2) \subseteq \alpha_k(v_2) = \alpha_k(v')$ . Moreover,  $v' <_{v_1}^{1,k} x'_1v_1w_1v_1z'_1 = v_2$  since  $|w_1| \geq k+1$  and  $\alpha_k(v_1w_1v_1) \subseteq \alpha_k(v_1)$ . Hence, for  $y' =_{\text{def}} x_2v'w_2v'z_2$  and  $y'' =_{\text{def}} x_2v_2w_2v'z_2$  we obtain  $y_1 = x_2v'z_2 <_{v'}^{1,k} y' <_{v_1}^{1,k} y'' <_{v_1}^{1,k} y_3$ .  $\square$

By definition,  $y \preceq_{\mathcal{M}}^{1,k} y'$  implies the existence of a  $<_{v_i}^{1,k}$  chain leading from  $y$  to  $y'$ . With help of Lemma 4.36 we show that one can choose this chain such that for the sequence of context words  $v_1, v_2, \dots, v_m$  it holds that  $|\alpha_k(v_1)| \geq |\alpha_k(v_2)| \geq \dots \geq |\alpha_k(v_m)|$ .

**Lemma 4.37.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $y, y' \in A^*$  with  $y \preceq_{\mathcal{M}}^{1,k} y'$ . Then there exist an  $m \geq 0$ ,  $y_0, \dots, y_m \in A^*$  and  $v_1, \dots, v_m \in W_{\mathcal{M}}^{1,k}$  with  $|\alpha_k(v_1)| \geq |\alpha_k(v_2)| \geq \dots \geq |\alpha_k(v_m)|$  and  $y = y_0 <_{v_1}^{1,k} y_1 <_{v_2}^{1,k} y_2 <_{v_3}^{1,k} \dots <_{v_m}^{1,k} y_m = y'$ .*

*Proof.* The proof is by contradiction, i.e., we assume that there exist  $y, y' \in A^*$  such that  $y \preceq_{\mathcal{M}}^{1,k} y'$  and  $y, y'$  do not have the property stated in the lemma. We choose a maximal



$m \geq 2$  such that there exist  $y_0, \dots, y_m \in A^*$  and  $v_1, \dots, v_m \in W_{\mathcal{M}}^{1,k}$  with  $y = y_0 \prec_{v_1}^{1,k} y_1 \prec_{v_2}^{1,k} y_2 \prec_{v_3}^{1,k} \dots \prec_{v_m}^{1,k} y_m = y'$ .

Define a *transposition* in a sequence of natural numbers  $n_1, \dots, n_l$  to be a pair of positions  $(i, j)$  with  $1 \leq i < j \leq l$  and  $n_i < n_j$ . Now we choose  $y_0, \dots, y_m \in A^*$  and  $v_1, \dots, v_m \in W_{\mathcal{M}}^{1,k}$  such that

- $y = y_0 \prec_{v_1}^{1,k} y_1 \prec_{v_2}^{1,k} y_2 \prec_{v_3}^{1,k} \dots \prec_{v_m}^{1,k} y_m = y'$  and
- the sequence  $|\alpha_k(v_1)|, |\alpha_k(v_2)|, \dots, |\alpha_k(v_m)|$  has a minimal number of transpositions.

By the assumption that  $y$  and  $y'$  disprove the lemma there exists some  $j$  with  $1 \leq j \leq m-1$  and  $|\alpha_k(v_j)| < |\alpha_k(v_{j+1})|$ . So we can apply Lemma 4.36 to the chain  $y_{j-1} \prec_{v_j}^{1,k} y_j \prec_{v_{j+1}}^{1,k} y_{j+1}$  and we obtain that at least one of the following statements holds.

1. There exists a  $y' \in A^*$  such that  $y_{j-1} \prec_{v_{j+1}}^{1,k} y' \prec_{v_j}^{1,k} y_{j+1}$ .
2. There exist  $y', y'' \in A^*$ ,  $v' \in W_{\mathcal{M}}^{1,k}$  with  $\alpha_k(v') = \alpha_k(v_{j+1})$  and  $y_{j-1} \prec_{v'}^{1,k} y' \prec_{v_j}^{1,k} y'' \prec_{v_j}^{1,k} y_{j+1}$ .

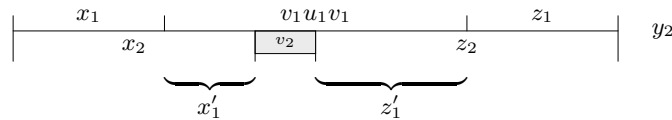
Statement 1 causes a contradiction to the choice of  $y_0, \dots, y_m, v_1, \dots, v_m$ , since the sequence  $|\alpha_k(v_1)|, \dots, |\alpha_k(v_{j-1})|, |\alpha_k(v_{j+1})|, |\alpha_k(v_j)|, |\alpha_k(v_{j+2})|, \dots, |\alpha_k(v_m)|$  has a smaller number of transpositions than the sequence  $|\alpha_k(v_1)|, |\alpha_k(v_1)|, \dots, |\alpha_k(v_m)|$ . If statement 2 holds then we obtain a contradiction to the maximal choice of  $m$ .  $\square$

By the preceding lemma, we can assume that the context words  $v_i$  in  $\prec_v^{1,k}$  chains are ordered by  $|\alpha_k(v_i)|$ . Our aim is to transform these chains into the form (4.12). So it remains to rearrange those parts of the chains whose context words  $v_i$  have the same  $|\alpha_k(v_i)|$ . The following lemma shows this for  $\prec_v^{1,k}$  chains of length two.

**Lemma 4.38.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$ ,  $y_1, y_2, y_3 \in A^*$  and  $v_1, v_2 \in W_{\mathcal{M}}^{1,k}$  such that  $|\alpha_k(v_1)| = |\alpha_k(v_2)|$  and  $y_1 \prec_{v_1}^{1,k} y_2 \prec_{v_2}^{1,k} y_3$ . Then  $y_1 \prec_{v_1}^{1,k} y_3$  or there exists a  $y'_2 \in A^*$  such that  $y_1 \prec_{v_2}^{1,k} y'_2 \prec_{v_1}^{1,k} y_3$ .*

*Proof.* By assumption,  $y_1 = x_1 v_1 z_1$ ,  $y_2 = x_1 v_1 u_1 v_1 z_1 = x_2 v_2 z_2$  and  $y_3 = x_2 v_2 u_2 v_2 z_2$  for suitable  $x_1, z_1, x_2, z_2 \in A^*$ ,  $u_1, u_2 \in A^{*\geq k+1}$  with  $\alpha_k(v_1 u_1 v_1) \subseteq \alpha_k(v_1)$  and  $\alpha_k(v_2 u_2 v_2) \subseteq \alpha_k(v_2)$ . If  $v_1 = v_2$  then we are done. So let us assume that  $v_1 \neq v_2$ . We compare the decompositions  $y_2 = x_1 v_1 u_1 v_1 z_1$  and  $y_2 = x_2 v_2 z_2$ .

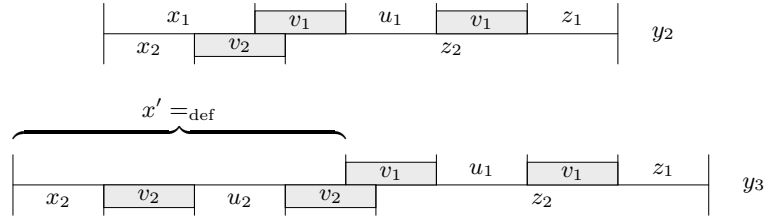
**Case 1:** The factor  $v_2$  of the decomposition  $y_2 = x_2 v_2 z_2$  is contained in the factor  $v_1 u_1 v_1$  of the decomposition  $y_2 = x_1 v_1 u_1 v_1 z_1$ . This means  $v_1 u_1 v_1 = x'_1 v_2 z'_1$ ,  $x_2 = x_1 x'_1$  and  $z_2 = z'_1 z_1$  for suitable words  $x'_1, z'_1 \in A^*$ .



From  $v_1 \neq v_2$  and Proposition 4.34 it follows that  $v_2$  is no factor of  $v_1$ . Therefore,  $v_1$  is both a prefix of  $x'_1 v_2$  and a suffix of  $v_2 z'_1$ . This shows that  $x'_1 v_2 u_2 v_2 z'_1 \in v_1 A^{*\geq k+1} v_1$ . Since

$v_1 \prec_{v_1}^{1,k} v_1 u_1 v_1 = x'_1 v_2 z'_1 \prec_{v_2}^{1,k} x'_1 v_2 u_2 v_2 z'_1$  we have  $v_1 \preceq_{\mathcal{M}}^{1,k} x'_1 v_2 u_2 v_2 z'_1$ . Now from Proposition 4.33.4 we get  $v_1 \prec_{v_1}^{1,k} x'_1 v_2 u_2 v_2 z'_1$ , and therefore  $y_1 = x_1 v_1 z_1 \prec_{v_1}^{1,k} x_1 x'_1 v_2 u_2 v_2 z'_1 z_1 = y_3$  by Proposition 1.19.

**Case 2:** The factor  $v_2$  of the decomposition  $y_2 = x_2 v_2 z_2$  is not contained in the factor  $v_1 u_1 v_1$  of the decomposition  $y_2 = x_1 v_1 u_1 v_1 z_1$ . From the assumption and Proposition 4.34 we obtain that  $v_1$  can not be a factor of  $v_2$ . Therefore, either  $|x_2| < |x_1|$  and  $v_2$  is a factor of  $x_1 v_1$ , or  $|z_2| < |z_1|$  and  $v_2$  is a factor of  $v_1 z_1$ . Without loss of generality we assume that  $|x_2| < |x_1|$  and that  $v_2$  is a factor of  $x_1 v_1$ .



Observe that  $x_1 v_1 \prec_{v_2}^{1,k} x' v_1$  where  $x'$  is defined as in the picture above. Moreover, since  $\alpha_k(v_1 u_1 v_1) \subseteq \alpha_k(v_1)$  we have  $v_1 z_1 \prec_{v_1}^{1,k} v_1 u_1 v_1 z_1$ . Together with Proposition 4.33.2 this implies  $y_1 = x_1 v_1 z_1 \prec_{v_2}^{1,k} x' v_1 z_1 \prec_{v_1}^{1,k} x' v_1 u_1 v_1 z_1 = y_3$ . Therefore, with  $y'_2 =_{\text{def}} x' v_1 z_1$  we obtain  $y_1 \prec_{v_2}^{1,k} y'_2 \prec_{v_1}^{1,k} y_3$ .  $\square$

As a consequence, for context words  $v_0$  and  $v_1$  with  $|\alpha_k(v_0)| = |\alpha_k(v_1)|$  we can swap successive closure operations.

**Corollary 4.39.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $v_0, v_1 \in W_{\mathcal{M}}^{1,k}$  with  $|\alpha_k(v_0)| = |\alpha_k(v_1)|$ . Then for every set of words  $L \subseteq A^*$  it holds that  $\langle\langle L \rangle_{\prec_{v_0}^{1,k}} \rangle_{\prec_{v_1}^{1,k}} = \langle\langle L \rangle_{\prec_{v_1}^{1,k}} \rangle_{\prec_{v_0}^{1,k}}$ .*

*Proof.* Since  $\langle\langle L \rangle_{\prec_{v_0}^{1,k}} \rangle_{\prec_{v_1}^{1,k}} = \bigcup_{y \in L} \langle\langle y \rangle_{\prec_{v_0}^{1,k}} \rangle_{\prec_{v_1}^{1,k}}$  it suffices to show the lemma for all  $L = \{y\}$  with  $y \in A^*$ . By symmetry, it is enough to show the inclusion  $\langle\langle y \rangle_{\prec_{v_0}^{1,k}} \rangle_{\prec_{v_1}^{1,k}} \subseteq \langle\langle y \rangle_{\prec_{v_1}^{1,k}} \rangle_{\prec_{v_0}^{1,k}}$ . This inclusion is an immediate consequence of the following claim.

**Claim.** *Let  $y = x_0 \prec_{v_0}^{1,k} x_1 \prec_{v_0}^{1,k} \dots \prec_{v_0}^{1,k} x_m = z_0 \prec_{v_1}^{1,k} z_1 \prec_{v_1}^{1,k} \dots \prec_{v_1}^{1,k} z_n = y'$  for  $m, n \geq 0$  and  $y, y', x_i, z_i \in A^*$ . Then there exist  $m', n' \geq 0$  and  $x'_i, z'_i \in A^*$  such that  $m' \leq m$ ,  $n' \leq n$  and  $y = z'_0 \prec_{v_1}^{1,k} z'_1 \prec_{v_1}^{1,k} \dots \prec_{v_1}^{1,k} z'_{n'} = x'_0 \prec_{v_0}^{1,k} x'_1 \prec_{v_0}^{1,k} \dots \prec_{v_0}^{1,k} x'_{m'} = y'$ .*

We show the claim by induction on  $m + n$  where the induction base consists of the cases  $m + n = 0, 1, 2$ . If  $m + n = 0$  or  $m + n = 1$  then  $m = 0$  or  $n = 0$  and we are done. If  $m + n = 2$  and  $m \neq n$  then again we are done. If  $m + n = 2$  and  $m = n$  then the claim follows from Lemma 4.38. This proves the induction base. So let us assume that there exists an  $l \geq 2$  such that our claim has been shown for all  $m, n \geq 0$  with  $m + n \leq l$ . Now we have to show it for all  $m, n \geq 0$  with  $m + n = l + 1$ .

If  $m = 0$  or  $n = 0$  then we are done. So we can assume that  $m, n \geq 1$ , and moreover, without loss of generality we may assume that  $n \geq m$  (the other case can be shown analogously). Note that  $n \geq m$  implies that  $n \geq 2$ . If we apply the induction hypothesis

to the chain  $y = x_0 \langle_{v_0}^{1,k} \cdots \langle_{v_0}^{1,k} x_m = z_0 \langle_{v_1}^{1,k} \cdots \langle_{v_1}^{1,k} z_{n-1}$  then we get  $\tilde{m}, \tilde{n} \geq 0$  and words  $\tilde{x}_i, \tilde{z}_i \in A^*$  such that  $\tilde{m} \leq m, \tilde{n} \leq n - 1$  and

$$y = \tilde{z}_0 \langle_{v_1}^{1,k} \tilde{z}_1 \langle_{v_1}^{1,k} \cdots \langle_{v_1}^{1,k} \tilde{z}_{\tilde{n}} = \tilde{x}_0 \langle_{v_0}^{1,k} \tilde{x}_1 \langle_{v_0}^{1,k} \cdots \langle_{v_0}^{1,k} \tilde{x}_{\tilde{m}} = z_{n-1}. \quad (4.13)$$

We apply the hypothesis again, this time to  $\tilde{x}_0 \langle_{v_0}^{1,k} \cdots \langle_{v_0}^{1,k} \tilde{x}_{\tilde{m}} = z_{n-1} \langle_{v_1}^{1,k} z_n = y'$  (this is possible since  $n \geq 2$  and therefore  $\tilde{m} + 1 \leq m + 1 \leq m + n - 1 = l$ ). We get  $m', \bar{n} \geq 0$  and words  $x'_i, \bar{z}_i \in A^*$  such that  $m' \leq \tilde{m}, \bar{n} \leq 1$  and

$$\tilde{x}_0 = \bar{z}_0 \langle_{v_1}^{1,k} \cdots \langle_{v_1}^{1,k} \bar{z}_{\bar{n}} = x'_0 \langle_{v_0}^{1,k} x'_1 \langle_{v_0}^{1,k} \cdots \langle_{v_0}^{1,k} x'_{m'} = y'. \quad (4.14)$$

If we merge the chains (4.13) and (4.14) at  $\tilde{x}_0$  then we obtain

$$y = \tilde{z}_0 \langle_{v_1}^{1,k} \cdots \langle_{v_1}^{1,k} \tilde{z}_{\tilde{n}} = \tilde{x}_0 = \bar{z}_0 \langle_{v_1}^{1,k} \cdots \langle_{v_1}^{1,k} \bar{z}_{\bar{n}} = x'_0 \langle_{v_0}^{1,k} \cdots \langle_{v_0}^{1,k} x'_{m'} = y'.$$

Let  $n' =_{\text{def}} \tilde{n} + \bar{n}$ ,  $z'_i =_{\text{def}} \tilde{z}_i$  for  $0 \leq i \leq \tilde{n}$ , and  $z'_{\tilde{n}+j} =_{\text{def}} \bar{z}_j$  for  $1 \leq j \leq \bar{n}$ . Since  $m' \leq \tilde{m} \leq m, \tilde{n} \leq n - 1$  and  $\bar{n} \leq 1$ , we obtain  $m' \leq m, n' \leq n$  and

$$y = z'_0 \langle_{v_1}^{1,k} z'_1 \langle_{v_1}^{1,k} \cdots \langle_{v_1}^{1,k} z'_{n'} = x'_0 \langle_{v_0}^{1,k} x'_1 \langle_{v_0}^{1,k} \cdots \langle_{v_0}^{1,k} x'_{m'} = y'.$$

This proves the claim.  $\square$

Now we can state the main theorem concerning the rearrangement of  $\langle_v^{1,k}$  chains: Every chain leading from  $y$  to  $y'$  can be transformed into a chain of the form (4.12). This means that the context words in this chain are ordered by the following rules: (i) context words  $v_i$  with a large  $|\alpha_k(v_i)|$  appear earlier than context words  $v_j$  with a small  $|\alpha_k(v_j)|$ , and (ii) all equal context words appear in one block (i.e., if there is some  $v_j$  between two appearances of  $v_i$  then  $v_i = v_j$ ).

**Theorem 4.40.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $y \in A^*$ . Then it holds that*

$$\langle y \rangle_{\succeq_{\mathcal{M}}^{1,k}} = \bigcup \langle \cdots \langle \langle y \rangle_{\langle_{v_0}^{1,k}} \rangle_{\langle_{v_1}^{1,k}} \rangle_{\langle_{v_2}^{1,k}} \cdots \rangle_{\langle_{v_m}^{1,k}} \rangle \quad (4.15)$$

where the union ranges over all  $m \geq 0$  and all pairwise different words  $v_0, \dots, v_m \in W_{\mathcal{M}}^{1,k}$  with  $|\alpha_k(v_0)| \geq |\alpha_k(v_1)| \geq \cdots \geq |\alpha_k(v_m)|$ .

*Proof.* First of all let us observe the following claim.

**Claim.** *Let  $m \geq 0$  and  $v_0, \dots, v_m \in W_{\mathcal{M}}^{1,k}$  with  $|\alpha_k(v_0)| = \cdots = |\alpha_k(v_m)|$  such that  $\{v_0, \dots, v_m\} = \{\bar{v}_0, \dots, \bar{v}_{\bar{m}}\}$  for  $\bar{m} \geq 0$  and pairwise different words  $\bar{v}_0, \dots, \bar{v}_{\bar{m}} \in W_{\mathcal{M}}^{1,k}$ . Then for all  $L \subseteq A^*$  it holds that  $\langle \cdots \langle \langle L \rangle_{\langle_{v_0}^{1,k}} \rangle_{\langle_{v_1}^{1,k}} \rangle_{\langle_{v_2}^{1,k}} \cdots \rangle_{\langle_{v_m}^{1,k}} \rangle = \langle \cdots \langle \langle L \rangle_{\langle_{\bar{v}_0}^{1,k}} \rangle_{\langle_{\bar{v}_1}^{1,k}} \rangle_{\langle_{\bar{v}_2}^{1,k}} \cdots \rangle_{\langle_{\bar{v}_{\bar{m}}}^{1,k}} \rangle$ .*

This is an easy consequence of Corollary 4.39 and the fact that  $\langle \langle L \rangle_{\langle_v^{1,k}} \rangle_{\langle_v^{1,k}} = \langle L \rangle_{\langle_v^{1,k}}$  for all  $v \in A^*$  and  $L \subseteq A^*$ .

We turn to the proof of the theorem. From the definition of  $\succeq_{\mathcal{M}}^{1,k}$  it is easy to see that in (4.15) the right-hand side is a subset of the left-hand side. In the remaining part of the proof we will show the reverse inclusion.

Trivially,  $y$  is an element of the right-hand side of equation (4.15). So let  $y' \in A^*$  with  $y \preceq_{\mathcal{M}}^{1,k} y'$  and  $y \neq y'$ . With Lemma 4.37 it follows that there exist an  $m \geq 1$ ,  $y_0, \dots, y_m \in A^*$  and  $v_1, \dots, v_m \in W_{\mathcal{M}}^{1,k}$  with  $|\alpha_k(v_1)| \geq |\alpha_k(v_2)| \geq \dots \geq |\alpha_k(v_m)|$  such that  $y = y_0 \prec_{v_1}^{1,k} y_1 \prec_{v_2}^{1,k} \dots \prec_{v_m}^{1,k} y_m = y'$ . We can subdivide this chain into maximal sections such that  $|\alpha_k(v_i)| = |\alpha_k(v_j)|$  for any  $i, j$  in this section. This means that there exist an  $l \geq 1$ , natural numbers  $n_0 > n_1 > \dots > n_{l-1}$  and positions  $0 = i_0 < i_1 < \dots < i_l = m$  such that  $|\alpha_k(v_{j'})| = n_j$  for all  $0 \leq j \leq l-1$  and  $i_j < j' \leq i_{j+1}$ . Hence, for each  $0 \leq j \leq l-1$  we have

1.  $y_{i_{j+1}} \in \langle \dots \langle \langle \langle y_{i_j} \rangle_{\prec_{v_{i_j+1}}^{1,k}} \rangle_{\prec_{v_{i_j+2}}^{1,k}} \rangle_{\prec_{v_{i_j+3}}^{1,k}} \dots \rangle_{\prec_{v_{i_j+1}}^{1,k}} \rangle$  and
2.  $n_j = |\alpha_k(v_{i_{j+1}})| = |\alpha_k(v_{i_j+2})| = \dots = |\alpha_k(v_{i_{j+1}})|$ .

For each  $0 \leq j \leq l-1$  we can choose some  $l_j \geq 0$  and suitable pairwise different words  $v_{j,0}, v_{j,1}, \dots, v_{j,l_j} \in W_{\mathcal{M}}^{1,k}$  such that  $\{v_{i_j+1}, v_{i_j+2}, \dots, v_{i_{j+1}}\} = \{v_{j,0}, v_{j,1}, \dots, v_{j,l_j}\}$ . So it holds that  $n_j = |\alpha_k(v_{j,0})| = \dots = |\alpha_k(v_{j,l_j})|$ . From our claim it follows that

$$y_{i_{j+1}} \in \langle \dots \langle \langle \langle y_{i_j} \rangle_{\prec_{v_{j,0}}^{1,k}} \rangle_{\prec_{v_{j,1}}^{1,k}} \rangle_{\prec_{v_{j,2}}^{1,k}} \dots \rangle_{\prec_{v_{j,l_j}}^{1,k}} \rangle \quad (4.16)$$

for  $0 \leq j \leq l-1$ . If we combine the facts (4.16) for  $0 \leq j \leq l-1$  then we obtain

$$y' = y_{i_l} \in \langle \dots \langle \langle \langle \langle \langle y \rangle_{\prec_{v_{0,0}}^{1,k}} \dots \rangle_{\prec_{v_{0,l_0}}^{1,k}} \rangle_{\prec_{v_{1,0}}^{1,k}} \dots \rangle_{\prec_{v_{1,l_1}}^{1,k}} \dots \rangle_{\prec_{v_{l-1,0}}^{1,k}} \dots \rangle_{\prec_{v_{l-1,l_{l-1}}}^{1,k}} \rangle \quad (4.17)$$

Now we have to show that all words  $v_{i,j}$  in (4.17) are pairwise different. For this purpose we assume that there exist words  $v_{i,j}$  and  $v_{i',j'}$  in (4.17) such that  $v_{i,j} = v_{i',j'}$ . If  $i \neq i'$  then  $\alpha_k(v_{i,j}) = n_i$ ,  $\alpha_k(v_{i',j'}) = n_{i'}$  and  $n_i \neq n_{i'}$ . This contradicts  $v_{i,j} = v_{i',j'}$ , and it follows that  $i = i'$ . Since by our claim the words  $v_{i,0}, v_{i,1}, \dots, v_{i,l_i}$  are pairwise different, we obtain  $j = j'$ . This shows that all words  $v_{i,j}$  in (4.17) are pairwise different. Therefore,  $y'$  is an element of the right-hand side of (4.15).  $\square$

The crucial point in Theorem 4.40 is that the context words  $v_0, \dots, v_m$  are pairwise different. This means that in (4.15) the union is finite and  $m$  is bounded. This allows us to prove the main theorem of this subsection.

**Theorem 4.41.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $y \in A^+$ . Then it holds that  $\langle y \rangle_{\preceq_{\mathcal{M}}^{1,0}} \in \mathcal{L}_{3/2}$  and  $\langle y \rangle_{\preceq_{\mathcal{M}}^{1,k}} \in \mathcal{B}_{3/2}$ .*

*Proof.* From Theorem 4.40 and Corollary 1.31 it follows that  $\langle y \rangle_{\preceq_{\mathcal{M}}^{1,k}} \in \tilde{\mathcal{B}}_k$  (here it is important that (i)  $W_{\mathcal{M}}^{1,k} \subseteq A^{*\geq k+1}$  and that (ii) the union in Theorem 4.40 is finite since  $W_{\mathcal{M}}^{1,k}$  is finite). So it suffices to show  $\tilde{\mathcal{B}}_0 \subseteq \mathcal{L}_{3/2}$  and  $\tilde{\mathcal{B}}_k \subseteq \mathcal{B}_{3/2}$ .

The inclusion  $\tilde{\mathcal{B}}_0 \subseteq \mathcal{L}_{3/2}$  is an immediate consequence of Definition 1.21 and Theorem 2.9. If we compare the Definitions 1.21 and 2.1 then we see that for the inclusion  $\tilde{\mathcal{B}}_k \subseteq \mathcal{B}_{3/2}$  it is enough to show that  $(\beta|\Gamma|\delta)_k \in \mathcal{B}_{3/2}$  for all  $\beta, \delta \in A^k$  and  $\Gamma \subseteq A^{k+1}$  (note that  $\{a\} \in \mathcal{B}_{3/2}$  for all  $a \in A$ ). It holds that  $(\beta|\Gamma|\delta)_k$  is even an element of  $\text{co}\mathcal{B}_{1/2}$  which can be seen as follows.

$$(\beta|\Gamma|\delta)_k = A^+ \setminus \left[ \underbrace{\left( \bigcup_{\gamma \in A^{k+1} \setminus \Gamma} A^* \gamma A^* \right)}_{\text{words of length } \geq k+1 \text{ containing a block } \gamma \notin \Gamma} \cup \underbrace{\left( \bigcup_{\substack{\beta' \in A^k \setminus \{\beta\} \\ \delta' \in A^k \setminus \{\delta\}}} \beta' A^+ \cup A^+ \delta' \right)}_{\text{words of length } \geq k+1 \text{ having a wrong prefix or suffix}} \cup \underbrace{\left( \bigcup_{w \in A^{+\leq k} \setminus (\beta|\Gamma|\delta)_k} \{w\} \right)}_{\text{words of length } \leq k \text{ that are not in } (\beta|\Gamma|\delta)_k} \right].$$

all nonempty words that are not in  $(\beta|\Gamma|\delta)_k$

□

### 4.3.3 Regular $\preceq_{\mathcal{M}}^{1,k}$ Co-Ideals are Finitely Generated

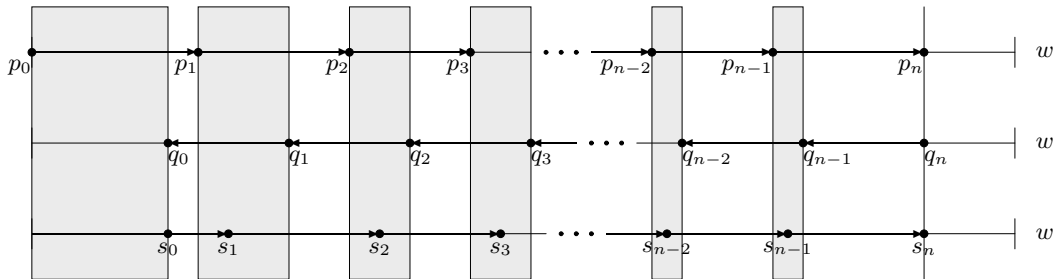
The main theorem of this subsection says that if the language accepted by some DFA  $\mathcal{M}$  is a  $\preceq_{\mathcal{M}}^{1,k}$  co-ideal then this co-ideal is finitely generated. There we will show that every sufficiently long word  $y' \in L(\mathcal{M})$  has a  $\preceq_{\mathcal{M}}^{1,k}$  predecessor in  $L(\mathcal{M})$ . For the proof it is necessary that in a sufficiently long word  $y$  we find ‘many’ appearances of the same context word  $v$  (see Lemma 4.45).

So we have to find context words, and therefore we start with a lemma which says: For a sufficiently large  $n$  and words  $u_1, u_2, \dots, u_n$  with  $\alpha_k(u_i) = \alpha_k(u_1 u_2 \dots u_n)$  we find a factor in  $u_1 u_2 \dots u_n$  which has nearly all properties of context words from  $W_{\mathcal{M}}^{1,k}$  (only the length is an exception). If we additionally bound the length of  $u_1 u_2 \dots u_n$  then it contains even a factor from  $W_{\mathcal{M}}^{1,k}$  (see Corollary 4.43).

**Lemma 4.42.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$ ,  $n =_{\text{def}} \mathcal{I}_{\mathcal{M}}$  and  $w, u_1, \dots, u_n \in A^{*\geq k+1}$  such that  $w = u_1 u_2 \dots u_n$  and  $\alpha_k(u_i) = \alpha_k(w)$  for  $1 \leq i \leq n$ . Then there exist  $w_1, w_2 \in A^*$  and  $r_1, \dots, r_n, l_1, \dots, l_n \in A^{*\geq k+1}$  such that the following holds.*

1.  $w = w_1 r_1 r_2 \dots r_n w_2 = w_1 l_n l_{n-1} \dots l_1 w_2$
2.  $\alpha_k(A^{-1} l_i) \subsetneq \alpha_k(w)$  and  $\alpha_k(r_i A^{-1}) \subsetneq \alpha_k(w)$  for  $1 \leq i \leq n$
3.  $\alpha_k(l_i) = \alpha_k(r_i) = \alpha_k(w)$  for  $1 \leq i \leq n$

*Proof.* Let  $m =_{\text{def}} |w|$  and choose suitable letters  $a_1, \dots, a_m$  such that  $w = a_1 a_2 \dots a_m$ . Next we describe a walk in the word  $w$  which has three stages and which is illustrated in the following picture.



In the first stage we start at position  $p_0 =_{\text{def}} 0$ , i.e., left to the first letter of  $w$  (cf. the first line in the picture). We walk to the *right* until we reach a position  $p_1 > p_0$  (i.e., the  $p_1$ -th letter of  $w$ ) such that  $\alpha_k(a_{p_0+1}a_{p_0+2}\cdots a_{p_1}) \subsetneq \alpha_k(w)$  and  $\alpha_k(a_{p_0+1}a_{p_0+2}\cdots a_{p_1}) = \alpha_k(w)$ . Now we continue our walk until we reach a position  $p_2 > p_1$  such that  $\alpha_k(a_{p_1+1}a_{p_1+2}\cdots a_{p_2}) \subsetneq \alpha_k(w)$  and  $\alpha_k(a_{p_1+1}a_{p_1+2}\cdots a_{p_2}) = \alpha_k(w)$ . In a similar way we obtain positions  $p_3, \dots, p_n$ .

In the second stage of the construction we start at position  $q_n =_{\text{def}} p_n$  and we walk to the *left* until we reach a position  $q_{n-1} < q_n$  such that  $\alpha_k(a_{q_{n-1}+2}a_{q_{n-1}+3}\cdots a_{q_n}) \subsetneq \alpha_k(w)$  and  $\alpha_k(a_{q_{n-1}+1}a_{q_{n-1}+2}\cdots a_{q_n}) = \alpha_k(w)$ . We continue the walk to the left until we reach a position  $q_{n-2} < q_{n-1}$  such that  $\alpha_k(a_{q_{n-2}+2}a_{q_{n-2}+3}\cdots a_{q_{n-1}}) \subsetneq \alpha_k(w)$  and  $\alpha_k(a_{q_{n-2}+1}a_{q_{n-2}+2}\cdots a_{q_{n-1}}) = \alpha_k(w)$ . If we continue this procedure we obtain the positions  $q_{n-3}, \dots, q_0$ .

The third stage is analogous to the first one, i.e., we walk to the *right*. Here we start at position  $s_0 =_{\text{def}} q_0$  and we obtain positions  $s_1, \dots, s_n$ .

First of all we make sure that the construction above is possible, i.e., during the construction we do not walk beyond the first and last letter of the word  $w$ . In the first stage it holds that  $p_i \leq |u_1u_2\cdots u_i|$  for  $0 \leq i \leq n$ , since  $\alpha_k(u_i) = \alpha_k(w)$  by assumption. This shows that we do not walk beyond the end of the word  $w$ . Hence, it holds that

$$0 = p_0 < p_1 < p_2 < \cdots < p_n \leq |w|. \quad (4.18)$$

The construction of the second stage is such that  $p_{n-1} \leq q_{n-1}$ , since otherwise the word  $a_{p_{n-1}+1}a_{p_{n-1}+2}\cdots a_{p_n}$  would be a factor of  $a_{q_{n-1}+2}a_{q_{n-1}+3}\cdots a_{q_n}$  and we would obtain  $\alpha_k(a_{p_{n-1}+1}a_{p_{n-1}+2}\cdots a_{p_n}) \subsetneq \alpha_k(w)$  (a contradiction to the construction in the first stage). Moreover, it holds that  $q_{n-2} < p_{n-1}$ , since otherwise the word  $a_{q_{n-2}+1}a_{q_{n-2}+2}\cdots a_{q_{n-1}}$  would be a factor of  $a_{p_{n-1}+1}a_{p_{n-1}+2}\cdots a_{p_{n-1}}$  and we would obtain  $\alpha_k(a_{q_{n-2}+1}a_{q_{n-2}+2}\cdots a_{q_{n-1}}) \subsetneq \alpha_k(w)$  (a contradiction to the construction in the second stage). Analogously one observes that  $q_i < p_{i+1} \leq q_{i+1}$  for  $0 \leq i \leq n-1$ . In particular we have  $p_1 \leq q_1$ . So if we walk from  $q_1$  to the left then at least at position  $p_0$  we reach a position such that  $a_{p_0+1}a_{p_0+2}\cdots a_{p_1}$  is a factor of  $a_{p_0+1}a_{p_0+2}\cdots a_{q_1}$  and therefore  $\alpha_k(a_{p_0+1}a_{p_0+2}\cdots a_{q_1}) = \alpha_k(w)$ . This implies that  $p_0 \leq q_0$ , i.e., during the second stage we do not walk beyond the beginning of the word  $w$ . So we have shown

$$0 = p_0 \leq q_0 < p_1 \leq q_1 < p_2 \leq q_2 < \cdots < p_{n-1} \leq q_{n-1} < p_n = q_n \leq |w|. \quad (4.19)$$

We use the same argumentation for the comparison of the second and the third stage and obtain

$$s_0 = q_0 < s_1 \leq q_1 < s_2 \leq q_2 < \cdots < s_{n-1} \leq q_{n-1} < s_n = q_n \leq |w|. \quad (4.20)$$

Now let us compare the first with the third stage. Immediately we obtain that  $p_i \leq s_i$  for  $0 \leq i \leq n$ , since we start the third stage at a position  $s_0$  and it holds that  $p_0 \leq q_0 = s_0$ . Together with (4.20) this implies  $s_n = q_n$ . We define the following words.

$$\begin{aligned}
w_1 &=_{\text{def}} a_1 a_2 \cdots a_{q_0} \\
w_2 &=_{\text{def}} a_{q_n+1} a_{q_n+2} \cdots a_{|w|} \\
r_i &=_{\text{def}} a_{s_{i-1}+1} a_{s_{i-1}+2} \cdots a_{s_i} & \text{for } 1 \leq i \leq n \\
l_i &=_{\text{def}} a_{q_{n-i}+1} a_{q_{n-i}+2} \cdots a_{q_{n-i+1}} & \text{for } 1 \leq i \leq n
\end{aligned}$$

By the construction process for the positions  $q_i$  and  $s_i$  it holds that  $\alpha_k(A^{-1}l_i) \subsetneq \alpha_k(w)$ ,  $\alpha_k(r_i A^{-1}) \subsetneq \alpha_k(w)$  and  $\alpha_k(l_i) = \alpha_k(r_i) = \alpha_k(w)$  for  $1 \leq i \leq n$ . From  $w \in A^{*\geq k+1}$  it follows that  $|\alpha_k(w)| \geq 1$  and therefore  $l_i, r_i \in A^{*\geq k+1}$  for  $1 \leq i \leq n$ . Since  $s_0 = q_0$  and  $s_n = q_n$ , we obtain  $w = w_1 r_1 r_2 \cdots r_n w_2 = w_1 l_n l_{n-1} \cdots l_1 w_2$ .  $\square$

**Corollary 4.43.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$ ,  $n =_{\text{def}} \mathcal{I}_{\mathcal{M}}$  and  $w, u_1, \dots, u_n \in A^{*\geq k+1}$  such that  $w = u_1 u_2 \cdots u_n$ ,  $|w| \leq \mathcal{C}_{\mathcal{M}}^k$  and  $\alpha_k(u_i) = \alpha_k(w)$  for  $1 \leq i \leq n$ . Then there exist  $w_1, w_2 \in A^*$  and  $v \in W_{\mathcal{M}}^{1,k}$  such that  $\alpha_k(v) = \alpha_k(w)$  and  $w = w_1 v w_2$ .*

*Proof.* From Lemma 4.42 we obtain words  $w_1, w_2 \in A^*$ ,  $r_1, \dots, r_n, l_1, \dots, l_n \in A^{*\geq k+1}$  such that  $w = w_1 r_1 r_2 \cdots r_n w_2 = w_1 l_n l_{n-1} \cdots l_1 w_2$ ,  $\alpha_k(A^{-1}l_i) \subsetneq \alpha_k(w)$ ,  $\alpha_k(r_i A^{-1}) \subsetneq \alpha_k(w)$  and  $\alpha_k(l_i) = \alpha_k(r_i) = \alpha_k(w)$ . Let  $v =_{\text{def}} r_1 r_2 \cdots r_n$  and observe that  $\alpha_k(v) = \alpha_k(w)$ . Now the corollary follows from Definition 4.31.  $\square$

In Corollary 4.43 we assumed a certain decomposition of the word  $w$ . Now we show a lemma which says that in every sufficiently long word  $y$  we find a factor  $w$  and a decomposition  $w = u_1 u_2 \cdots u_m$  with  $|w| \leq \mathcal{C}_{\mathcal{M}}^k$  and  $\alpha_k(u_i) = \alpha_k(w)$ . In this lemma,  $m$  will be substantial larger than the number  $n = \mathcal{I}_{\mathcal{M}}$  from Corollary 4.43. Moreover, the number  $m$  depends on the length of the words  $u_i$  (which will be all of the same length), and this length depends on  $l =_{\text{def}} |\alpha_k(u_i)|$ . This is important owing to the following considerations which will be exploit in the proof of Lemma 4.45.

- The dependence of  $m$  on  $|u_i|$  ensures that  $m$  is substantial larger than  $|A|^{\mathcal{I}_{\mathcal{M}} \cdot |u_i|}$ . Therefore, there exists a block  $b \in A^*$  such that  $|b| = \mathcal{I}_{\mathcal{M}} \cdot |u_i|$  and  $b$  appears at several positions in  $w = u_1 u_2 \cdots u_m$  (i.e.,  $b = u_{j+1} u_{j+2} \cdots u_{j+\mathcal{I}_{\mathcal{M}}}$  for several indices  $j$ ).
- The dependence of  $|u_i|$  on  $l = |\alpha_k(u_i)|$  ensures that for each block  $b$  it holds that  $|b| \leq \mathcal{C}_{\mathcal{M}}^k$ . This makes Corollary 4.43 applicable to the blocks  $b$ .

**Lemma 4.44.** *Let  $\mathcal{M}$  be a DFA,  $k \geq 0$  and  $y \in A^*$ . If  $|y| \geq \mathcal{F}_{\mathcal{M}}^k(|\alpha_k(y)|)$  then there exists an  $l$  with  $1 \leq l \leq |\alpha_k(y)|$  and a decomposition  $y = y_1 u_1 u_2 \cdots u_m y_2$  with  $y_1, y_2 \in A^*$ ,  $m = (\mathcal{I}_{\mathcal{M}} + 1) \cdot (|\mathcal{M}| \cdot |A|^{\mathcal{I}_{\mathcal{M}} \cdot \mathcal{F}_{\mathcal{M}}^k(l-1)+1})$  and  $|u_i| = \mathcal{F}_{\mathcal{M}}^k(l-1)$  for  $1 \leq i \leq m$  such that*

1.  $|\alpha_k(u_1 u_2 \cdots u_m)| = l$  and
2.  $\alpha_k(u_i) = \alpha_k(u_1 u_2 \cdots u_m)$  for  $1 \leq i \leq m$ .

*Proof.* We show the lemma by induction on  $|\alpha_k(y)| \geq 0$ . The induction base (i.e., the case  $|\alpha_k(y)| = 0$ ) holds trivially, since there are no words  $y \in A^*$  with  $|\alpha_k(y)| = 0$  and  $|y| \geq \mathcal{F}_{\mathcal{M}}^k(0) = k + 1$ .

Now we assume that there is some  $r \geq 0$  such that the lemma has been shown for all  $y \in A^*$  with  $|\alpha_k(y)| \leq r$ . Based on this assumption we will show the lemma for all  $y \in A^*$  with  $|\alpha_k(y)| = r + 1$ . Let  $m' =_{\text{def}} (\mathcal{I}_{\mathcal{M}} + 1) \cdot (|\mathcal{M}| \cdot |A|^{\mathcal{I}_{\mathcal{M}} \cdot \mathcal{F}_{\mathcal{M}}^k(r)+1})$  and let  $y \in A^*$  with

$|\alpha_k(y)| = r + 1$  and  $|y| \geq \mathcal{F}_{\mathcal{M}}^k(r + 1)$ . By definition this means that  $|y| \geq \mathcal{F}_{\mathcal{M}}^k(r) \cdot m'$ . Hence,  $y$  can be written as  $y = y_1 u_1 u_2 \cdots u_{m'} y_2$  for suitable  $y_1, y_2 \in A^*$  and factors  $u_1, \dots, u_{m'}$  of length  $\mathcal{F}_{\mathcal{M}}^k(r)$ .

If  $\alpha_k(u_i) = \alpha_k(y)$  for  $1 \leq i \leq m'$  then we define  $l =_{\text{def}} r + 1$ ,  $m =_{\text{def}} m'$  and we are done. Otherwise there exists some  $j$  with  $1 \leq j \leq m'$  and  $|\alpha_k(u_j)| \leq r$ . So we can apply the induction hypothesis to  $u_j$ . Note that the word  $u_j$  satisfies  $|u_j| = \mathcal{F}_{\mathcal{M}}^k(r) \geq \mathcal{F}_{\mathcal{M}}^k(|\alpha_k(u_j)|)$ . Therefore, from the hypothesis we get an  $l \geq 1$  and a decomposition  $u_j = u'_1 u'_1 u'_2 \cdots u'_m u'_2$  with  $m = (\mathcal{I}_{\mathcal{M}} + 1) \cdot |A|^{\mathcal{I}_{\mathcal{M}} \cdot \mathcal{F}_{\mathcal{M}}^k(l-1) + 1}$  and  $|u'_i| = \mathcal{F}_{\mathcal{M}}^k(l - 1)$  for  $1 \leq i \leq m$  such that  $|\alpha_k(u'_1 u'_2 \cdots u'_m)| = l$  and  $\alpha_k(u'_i) = \alpha_k(u'_1 u'_2 \cdots u'_m)$  for  $1 \leq i \leq m$ . Let  $\bar{y}_1 =_{\text{def}} u_1 \cdots u_{j-1} u'_1$  and  $\bar{y}_2 =_{\text{def}} u'_2 u_{j+1} \cdots u_{m'}$ . We obtain  $y = \bar{y}_1 u'_1 u'_2 \cdots u'_m \bar{y}_2$  and  $1 \leq l = |\alpha_k(u'_1)| \leq |\alpha_k(y)|$ . This completes the induction step.  $\square$

Next we show that for every sufficiently long word  $y$  we find a context word  $v$  that appears several times as a factor in  $y$ .

**Lemma 4.45.** *Let  $\mathcal{M}$  be a DFA,  $r =_{\text{def}} |\mathcal{M}| + 1$  and  $k \geq 0$ . Every  $y \in A^*$  with  $|y| \geq \mathcal{C}_{\mathcal{M}}^k$  can be written as  $y = x v y_1 v y_2 \cdots y_{r-1} v z$  with  $x, z \in A^*$ ,  $y_i \in A^{\geq k+1}$ ,  $v \in \mathbb{W}_{\mathcal{M}}^{1,k}$  and  $\alpha_k(v y_1 v y_2 v \cdots y_{r-1} v) = \alpha_k(v)$ .*

*Proof.* Observe that  $|y| \geq \mathcal{C}_{\mathcal{M}}^k$  implies that  $|y| \geq \mathcal{F}_{\mathcal{M}}^k(|\alpha_k(y)|)$ . So we can apply Lemma 4.44 to  $y$ . We get natural numbers  $l, m$  and words  $y_1, y_2, u_1, \dots, u_m \in A^*$  having the properties stated in Lemma 4.44. Note that in particular this implies  $u_i \in A^{\geq k+1}$  for  $1 \leq i \leq m$ .

Let  $m' =_{\text{def}} |\mathcal{M}| \cdot |A|^{\mathcal{I}_{\mathcal{M}} \cdot \mathcal{F}_{\mathcal{M}}^k(l-1) + 1}$ . Since  $m = (\mathcal{I}_{\mathcal{M}} + 1) \cdot m'$  we can divide the sequence  $u_1, u_2, \dots, u_m$  into  $m'$  blocks of  $\mathcal{I}_{\mathcal{M}} + 1$  elements. For this let  $i_j =_{\text{def}} j \cdot (\mathcal{I}_{\mathcal{M}} + 1)$  for  $0 \leq j \leq m'$ . Therefore, if we define  $w_j =_{\text{def}} u_{i_{j-1}+1} u_{i_{j-1}+2} \cdots u_{i_j}$  for  $1 \leq j \leq m'$  then we obtain

$$y = y_1 \cdot \underbrace{w_1 u_{i_1} w_2 u_{i_2} w_3 u_{i_3} \cdots w_{m'} u_{i_{m'}}}_{w =_{\text{def}}} \cdot y_2.$$

Since  $\alpha_k(u_i) = \alpha_k(w)$  for  $1 \leq i \leq m$  we have  $\alpha_k(u_i) = \alpha_k(w_j)$  for  $1 \leq i \leq m$  and  $1 \leq j \leq m'$ . Moreover, from  $|u_i| = \mathcal{F}_{\mathcal{M}}^k(l - 1)$  for  $1 \leq i \leq m$  it follows that  $|w_j| = \mathcal{I}_{\mathcal{M}} \cdot \mathcal{F}_{\mathcal{M}}^k(l - 1) \leq \mathcal{F}_{\mathcal{M}}^k(l) \leq \mathcal{C}_{\mathcal{M}}^k$  for  $1 \leq j \leq m'$ . Therefore, we can apply Corollary 4.43 to the words  $w_j = u_{i_{j-1}+1} u_{i_{j-1}+2} \cdots u_{i_j}$ . For  $1 \leq j \leq m'$  we obtain words  $w_{j,1}, w_{j,2} \in A^*$  and  $v_j \in \mathbb{W}_{\mathcal{M}}^{1,k}$  such that  $\alpha_k(v_j) = \alpha_k(w_j) = \alpha_k(w)$  and  $w_j = w_{j,1} v_j w_{j,2}$ . This yields the following decomposition for  $y$ .

$$y = y_1 \cdot w_{1,1} v_1 w_{1,2} u_{i_1} w_{2,1} v_2 w_{2,2} u_{i_2} w_{3,1} v_3 w_{3,2} u_{i_3} \cdots w_{m',1} v_{m'} w_{m',2} u_{i_{m'}} \cdot y_2. \quad (4.21)$$

Observe that  $|v_j| \leq |w_j| = \mathcal{I}_{\mathcal{M}} \cdot \mathcal{F}_{\mathcal{M}}^k(l - 1)$  for  $1 \leq j \leq m'$ . On the one hand in (4.21) we have  $m'$  words  $v_j \in \mathbb{W}_{\mathcal{M}}^{1,k}$  with  $|v_j| \leq \mathcal{I}_{\mathcal{M}} \cdot \mathcal{F}_{\mathcal{M}}^k(l - 1)$ . On the other hand the number of words from  $A^*$  with length  $\leq \mathcal{I}_{\mathcal{M}} \cdot \mathcal{F}_{\mathcal{M}}^k(l - 1)$  is less than  $|A|^{\mathcal{I}_{\mathcal{M}} \cdot \mathcal{F}_{\mathcal{M}}^k(l-1) + 1}$ . Therefore, at least one of these words, say the word  $v$ , appears more than  $m' / (|A|^{\mathcal{I}_{\mathcal{M}} \cdot \mathcal{F}_{\mathcal{M}}^k(l-1) + 1}) = |\mathcal{M}|$  times in the decomposition (4.21). Hence there exist words  $x, z \in A^*$  and  $y_1, \dots, y_{r-1} \in A^{\geq k+1}$  such that  $y = x v y_1 v y_2 v \cdots y_{r-1} v z$ . Moreover, we have  $\alpha_k(v y_1 v y_2 v \cdots y_{r-1} v) = \alpha_k(w)$ , since  $v y_1 v y_2 v \cdots y_{r-1} v$  is a factor of  $w$  and since  $\alpha_k(v) = \alpha_k(w)$ . This proves the lemma.  $\square$



The main result of this section shows that if the language accepted by some DFA  $\mathcal{M}$  is a  $\preceq_{\mathcal{M}}^{1,k}$  co-ideal then this co-ideal is finitely generated. Together with Theorem 4.41 this implies that  $L(\mathcal{M}) \in \mathcal{B}_{3/2}$  (and if  $k = 0$  then  $L(\mathcal{M}) \in \mathcal{L}_{3/2}$ ).

**Theorem 4.46.** *Let  $\mathcal{M}$  be a DFA and  $k \geq 0$ . If  $L(\mathcal{M})$  is a  $\preceq_{\mathcal{M}}^{1,k}$  co-ideal then it is even finitely generated (i.e., there exists a finite set  $D \subseteq A^+$  with  $\langle D \rangle_{\preceq_{\mathcal{M}}^{1,k}} = L(\mathcal{M})$ ).*

*Proof.* Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA,  $k \geq 0$  and assume that  $L(\mathcal{M})$  is a  $\preceq_{\mathcal{M}}^{1,k}$  co-ideal. It suffices to show that for every  $y \in L(\mathcal{M})$  with  $|y| \geq \mathcal{C}_{\mathcal{M}}^k$  there exists a  $y' \in L(\mathcal{M})$  with  $|y'| < \mathcal{C}_{\mathcal{M}}^k$  and  $y' \preceq_{\mathcal{M}}^{1,k} y$ . To see this it is enough to show the following claim.

**Claim.** *For every  $y \in L(\mathcal{M})$  with  $|y| \geq \mathcal{C}_{\mathcal{M}}^k$  there exists a  $y' \in L(\mathcal{M})$  with  $|y'| < |y|$  and  $y' \preceq_{\mathcal{M}}^{1,k} y$ .*

Let  $y \in L(\mathcal{M})$  with  $|y| \geq \mathcal{C}_{\mathcal{M}}^k$  and let  $r =_{\text{def}} |\mathcal{M}| + 1$ . By Lemma 4.45,  $y$  can be written as  $y = xvy_1vy_2v \cdots y_{r-1}vz$  for words  $x, z \in A^*$ ,  $y_i \in A^{\geq k+1}$  and  $v \in W_{\mathcal{M}}^{1,k}$  such that  $\alpha_k(vy_1vy_2v \cdots y_{r-1}v) = \alpha_k(v)$ . For  $1 \leq i \leq r$  let  $\tilde{s}_i =_{\text{def}} \delta(s_0, xvy_1vy_2 \cdots vy_{i-1})$ . Since  $r > |\mathcal{M}|$  there exist positions  $j, j'$  with  $1 \leq j < j' \leq r$  such that  $\tilde{s}_j = \tilde{s}_{j'}$ . It follows that in  $y$  we can cut the factor  $vy_jvy_{j+1} \cdots vy_{j'-1}$  without leaving the language. This means that the word  $y' =_{\text{def}} xvy_1vy_2 \cdots vy_{j-1} \cdot vy_jvy_{j'+1} \cdots vy_{r-1}vz$  is still an element of  $L(\mathcal{M})$ . Since  $|y'| < |y|$  it remains to show that  $y' \preceq_{\mathcal{M}}^{1,k} y$ .

Let  $x' =_{\text{def}} xvy_1 \cdots vy_{j-1}$ ,  $z' =_{\text{def}} y_jv \cdots y_{r-1}vz$ , and  $w =_{\text{def}} y_jvy_{j+1} \cdots vy_{j'-1}$ . Hence we have  $y' = x'vz'$  and  $y = x'vwvz'$ . Since  $vwv$  is a factor of  $vy_1vy_2v \cdots y_{r-1}v$  it follows that  $\alpha_k(vwv) = \alpha_k(v)$ . Moreover, we have  $w \in A^{\geq k+1}$  since  $w$  contains at least the factor  $y_j \in A^{\geq k+1}$ . This shows  $y' <_v^{1,k} y$  and it follows that  $y' \preceq_{\mathcal{M}}^{1,k} y$ .  $\square$

In view of Theorem 4.8, which shows that  $(A^*, \preceq_{\mathcal{M}}^{0,k})$  is a wpos, one might ask if  $(A^*, \preceq_{\mathcal{M}}^{1,k})$  is also a wpos. Unfortunately this does not hold. To see this it suffices to construct an infinite set of pairwise incomparable words. Choose two arbitrary letters  $a$  and  $b$  from the alphabet. Let  $n =_{\text{def}} \mathcal{I}_{\mathcal{M}}$ ,  $m =_{\text{def}} \mathcal{C}_{\mathcal{M}}^k$  and  $w_i =_{\text{def}} (a^m b^m)^i$  for  $i \geq 1$ . We will see that  $w_i \not\preceq_{\mathcal{M}}^{1,k} w_j$  for all  $i \neq j$ .

Assume that we find a context word  $v \in W_{\mathcal{M}}^{1,k}$  in some  $w_i$ . We want to see that  $v$  is of the form  $v = a^l$  or  $v = b^l$  for some  $l$ . From the definition of context words, i.e., Definition 4.31, it follows that  $v = r_1 \cdots r_n$  for suitable  $r_j \in A^{\geq k+1}$  with  $\alpha_k(r_j) = \alpha_k(v)$ . From  $v \in W_{\mathcal{M}}^{1,k}$  it follows that  $|v| \leq m$ . So we obtain four possibilities: (i)  $v$  is a factor of  $a^m$ , (ii)  $v$  is a factor of  $b^m$ , (iii)  $v$  is a factor of  $a^m b^m$  and is neither a factor of  $a^m$  nor a factor of  $b^m$ , (iv)  $v$  is a factor of  $b^m a^m$  and is neither a factor of  $a^m$  nor a factor of  $b^m$ . In the cases (i) and (ii) we are done. Since case (iv) is analogous to case (iii), it suffices to consider case (iii). In this case the set  $\alpha_k(v)$  contains at least one word from  $A^* a A^*$  and at least one word from  $A^* b A^*$ . Since  $\alpha_k(v) = \alpha_k(r_1) = \alpha_k(r_n)$  there appears a letter  $b$  in  $r_1$  and there appears a letter  $a$  in  $r_2$ . This contradicts the assumption that  $v = r_1 \cdots r_n$  is a factor of  $a^m b^m$ . Therefore, in each  $w_i$  we find only context words  $v \in W_{\mathcal{M}}^{1,k}$  that are of the form  $v = a^l$  or  $v = b^l$ .

For  $\preceq_{\mathcal{M}}^{1,k}$  extensions (applied to the words  $w_i$ ) this has the following consequence: These extensions can only insert letters  $a$  into  $a$ -blocks and letters  $b$  into  $b$ -blocks. These exten-

sions cannot change the number of  $a$ -blocks and  $b$ -blocks in the words  $w_i$ . Since different words  $w_i, w_j$  have different such numbers we obtain that  $w_i \not\preceq_{\mathcal{M}}^{1,k} w_j$  for  $i \neq j$ .

#### 4.3.4 Languages from $\mathcal{FP}(\mathbb{P}_1^c)$ and $\mathcal{FP}(\mathbb{P}_1^b)$ are Regular $\preceq_{\mathcal{M}}^{1,k}$ Co-Ideals

We already know that (i) the  $\preceq_{\mathcal{M}}^{1,k}$  upward closure of a nonempty word is in  $\mathcal{B}_{3/2}$  and (ii) if the language accepted by some DFA  $\mathcal{M}$  is a  $\preceq_{\mathcal{M}}^{1,k}$  co-ideal then it is finitely generated. It follows that if  $L(\mathcal{M})$  is a  $\preceq_{\mathcal{M}}^{1,k}$  co-ideal then  $L(\mathcal{M}) \in \mathcal{B}_{3/2}$ . In this section we establish the missing connection to the forbidden-pattern class  $\mathcal{FP}(\mathbb{P}_1^b)$ . We show that if the language accepted by the DFA  $\mathcal{M}$  is in  $\mathcal{FP}(\mathbb{P}_1^b)$  then  $L(\mathcal{M})$  is a  $\preceq_{\mathcal{M}}^{1,k}$  co-ideal for a suitable  $k \geq 0$ . For the forbidden-pattern class  $\mathcal{FP}(\mathbb{P}_1^c)$  we obtain an analogous result.

So we have to connect patterns from  $\mathbb{P}_1^b$  on the one hand with the word extensions  $\preceq_{\mathcal{M}}^{1,k}$  on the other hand. This connection is not established directly, but we introduce an intermediate pattern. In Lemma 4.48 we show that the absence of patterns from  $\mathbb{P}_1^b$  implies the absence of this intermediate pattern (Lemma 4.49 states the analogous result for  $\mathbb{P}_1^c$ ). Then in the proof of Theorem 4.50 we show that the absence of the intermediate pattern in some DFA  $\mathcal{M}$  implies that  $L(\mathcal{M})$  is closed under  $\preceq_{\mathcal{M}}^{1,k}$  for a suitable  $k$ . The analogous result for  $\mathcal{FP}(\mathbb{P}_1^c)$  is given in Theorem 4.51.

We start with a proposition that provides a way to rewrite the intermediate pattern.

**Proposition 4.47.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA and  $k =_{\text{def}} 3 \cdot \mathcal{I}_{\mathcal{M}}$ . Suppose that there exist states  $s_1, s_2$  and words  $x, z \in A^*$ ,  $v, w \in A^{*\geq k+1}$  with  $\alpha_k(vwv) \subseteq \alpha_k(vv)$ ,  $s_0 \xrightarrow{x} s_1 \xrightarrow{v} s_1 \xrightarrow{z} +$  and  $s_1 \xrightarrow{w} s_2 \xrightarrow{v} s_2 \xrightarrow{z} -$ . Then we can choose such  $s_1, s_2, x, z, v, w$  that additionally satisfy  $v = v'u$  and  $w = w'u$  for  $v', w' \in A^{*\geq k+1}$ ,  $u \in A^{+\leq \mathcal{I}_{\mathcal{M}}}$  with  $\delta^{uu} = \delta^u$ .*

*Proof.* We define below witnessing states  $\tilde{s}_1, \tilde{s}_2$  and witnessing words  $\tilde{x}, \tilde{z}, \tilde{v}, \tilde{w}$  having the required properties. By Corollary 1.17,  $v$  can be written as  $v = w_1 u w_2$  for words  $w_1, w_2 \in A^*$  and  $u \in A^{+\leq \mathcal{I}_{\mathcal{M}}}$  with  $\delta^{uu} = \delta^u$ . We make the following definitions.

$$\begin{aligned} \tilde{x} &=_{\text{def}} x w_1 u & \tilde{v} &=_{\text{def}} w_2 v v w_1 u & \tilde{s}_1 &=_{\text{def}} \delta(s_1, w_1 u) \\ \tilde{z} &=_{\text{def}} w_2 z & \tilde{w} &=_{\text{def}} w_2 w w_1 u & \tilde{s}_2 &=_{\text{def}} \delta(s_2, w_1 u) \end{aligned}$$

With  $v' =_{\text{def}} w_2 v v w_1$  and  $w' =_{\text{def}} w_2 w w_1$  we get  $v', w' \in A^{*\geq k+1}$ ,  $\tilde{v} = v'u$  and  $\tilde{w} = w'u$ . In particular this shows  $|\tilde{w}| \geq k+1$  and  $|\tilde{v}| \geq k+1$ . The following facts can be easily observed.

$$\begin{aligned} \delta(\tilde{s}_1, \tilde{v}) &= \delta(s_1, w_1 u \tilde{v}) = \delta(s_1, w_1 u w_2 v v w_1 u) = \delta(s_1, v^3 w_1 u) = \delta(s_1, w_1 u) = \tilde{s}_1 \\ \delta(\tilde{s}_2, \tilde{v}) &= \delta(s_2, w_1 u \tilde{v}) = \delta(s_2, w_1 u w_2 v v w_1 u) = \delta(s_2, v^3 w_1 u) = \delta(s_2, w_1 u) = \tilde{s}_2 \\ \delta(\tilde{s}_1, \tilde{w}) &= \delta(s_1, w_1 u \tilde{w}) = \delta(s_1, w_1 u w_2 w w_1 u) = \delta(s_1, v w w_1 u) = \delta(s_2, w_1 u) = \tilde{s}_2 \\ \delta(\tilde{s}_1, \tilde{z}) &= \delta(s_1, w_1 u \tilde{z}) = \delta(s_1, w_1 u w_2 z) = \delta(s_1, v z) = \delta(s_1, z) \in S' \\ \delta(\tilde{s}_2, \tilde{z}) &= \delta(s_2, w_1 u \tilde{z}) = \delta(s_2, w_1 u w_2 z) = \delta(s_2, v z) = \delta(s_2, z) \notin S' \end{aligned}$$

Hence we obtain  $s_0 \xrightarrow{\tilde{x}} \tilde{s}_1 \xrightarrow{\tilde{v}} \tilde{s}_1 \xrightarrow{\tilde{z}} +$  and  $\tilde{s}_1 \xrightarrow{\tilde{w}} \tilde{s}_2 \xrightarrow{\tilde{v}} \tilde{s}_2 \xrightarrow{\tilde{z}} -$ . Since  $\tilde{v} \tilde{w} \tilde{v}$  is a factor of  $v v v v v v v v$  we get  $\alpha_k(\tilde{v} \tilde{w} \tilde{v}) \subseteq \alpha_k(v v v v v v v v) \subseteq \alpha_k(vv)$  (for the latter inclusion we use the facts  $\alpha_k(vwv) \subseteq \alpha_k(vv)$  and  $|v| \geq k+1$ ). Since  $vv$  is a factor of  $\tilde{v}$  it holds that  $\alpha_k(vv) \subseteq \alpha_k(\tilde{v})$  and therefore  $\alpha_k(\tilde{v} \tilde{w} \tilde{v}) \subseteq \alpha_k(\tilde{v} \tilde{v})$ .  $\square$

**Lemma 4.48.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA and  $k =_{\text{def}} 3 \cdot \mathcal{I}_{\mathcal{M}}$ . Suppose that there exist states  $s_1, s_2$  and words  $x, z \in A^*$ ,  $v, w \in A^{*\geq k+1}$  such that  $\alpha_k(vwv) \subseteq \alpha_k(vv)$ ,  $s_0 \xrightarrow{x} s_1 \xrightarrow{v} s_1 \xrightarrow{z} +$  and  $s_1 \xrightarrow{w} s_2 \xrightarrow{v} s_2 \xrightarrow{z} -$ . Then  $\mathcal{M}$  has a pattern from  $\mathbb{P}_1^{\mathcal{B}}$ .*

*Proof.* By Proposition 4.47 we may assume that  $v$  and  $w$  are of the form  $v = v'u$  and  $w = w'u$  for  $v', w' \in A^{*\geq k+1}$ ,  $u \in A^{+\leq \mathcal{I}_{\mathcal{M}}}$  with  $\delta^{uu} = \delta^u$ . It follows that both states  $s_1$  and  $s_2$  have a  $u$ -loop. From Corollary 1.15 we obtain a decomposition  $w' = w'_0 u_1 w'_1 \cdots u_m w'_m$  such that  $w'_i, u_i \in A^{+\leq \mathcal{I}_{\mathcal{M}}}$  and  $\delta^{u_i u_i} = \delta^{u_i}$ . Let  $u_0 =_{\text{def}} u$  and  $u_{m+1} =_{\text{def}} u$ . Then it holds that  $u_0, w'_0, u_1, \dots, w'_m, u_{m+1} \in A^{+\leq \mathcal{I}_{\mathcal{M}}}$  and  $\delta^{u_i u_i} = \delta^{u_i}$  for  $0 \leq i \leq m+1$ .

Next we want to observe that the factors  $u_i w'_i u_{i+1}$  appear as factors in a loop at  $s_1$  and in a loop at  $s_2$ . Note that  $|u_i w'_i u_{i+1}| \leq 3 \cdot \mathcal{I}_{\mathcal{M}} < k+1$  for  $0 \leq i \leq m$  and that  $u_0 w'_0 u_1 \cdots w'_m u_{m+1} = w'w'u = uw$  is a factor of  $vw$ . Hence,  $u_i w'_i u_{i+1}$  appears as a factor in some element of  $\alpha_k(vwv)$  for  $0 \leq i \leq m$ . From  $\alpha_k(vwv) \subseteq \alpha_k(vv)$  it follows that  $u_i w'_i u_{i+1}$  is a factor of  $vv$  for  $0 \leq i \leq m$ . This means that for  $0 \leq i \leq m$  there exist  $v'_i, v''_i$  with  $vv = v'_i u_i w'_i u_{i+1} v''_i$ .

In order to show that  $\mathcal{M}$  has a pattern from  $\mathbb{P}_1^{\mathcal{B}}$  we let  $m' =_{\text{def}} m+1$  and we make the following definitions.

$$\begin{aligned} w_0 &=_{\text{def}} u_0 & p_0 &=_{\text{def}} (u_0, v'_0 u_0) \\ w_i &=_{\text{def}} w'_{i-1} u_i \quad \text{for } 1 \leq i \leq m' & p_i &=_{\text{def}} (u_i, v''_{i-1} v'_i u_i) \quad \text{for } 1 \leq i \leq m' - 1 \\ & & p_{m'} &=_{\text{def}} (u_{m'}, v''_{m'-1} u_{m'}) \end{aligned}$$

Note that  $w_i$  and both components of  $p_i$  are elements of  $A^+$  for  $0 \leq i \leq m'$ . From Definition 3.36 it follows that  $p_i \in \mathcal{B}$  for  $0 \leq i \leq m'$ . Hence, in view of Definition 3.6 we can refer to the first (respectively, second) component of  $p_i$  as  $\overline{p}_i^\circ$  (respectively,  $\overline{p}_i$ ) for  $0 \leq i \leq m'$ . Let  $p =_{\text{def}} (w_0, p_0, \dots, w_{m'}, p_{m'})$  and note that  $p \in \mathbb{P}_1^{\mathcal{B}}$  by Definition 3.4. In order to show that  $\mathcal{M}$  has a pattern from  $\mathbb{P}_1^{\mathcal{B}}$  (see Definition 3.5) it is enough to see that  $s_1 \xrightarrow{p} s_2$  (since we already know that  $s_0 \xrightarrow{z} s_1 \xrightarrow{z} +$  and  $s_2 \xrightarrow{z} -$ ).

For  $0 \leq i \leq m'$  let  $q_i =_{\text{def}} \delta(s_1, w_0 \overline{p}_0 \cdots w_{i-1} \overline{p}_{i-1} w_i)$  and  $r_i =_{\text{def}} \delta(q_i, \overline{p}_i)$ . Observe that  $w_i$  and  $\overline{p}_i$  lead to a  $\overline{p}_i^\circ$ -loop in  $\mathcal{M}$  since both words have the suffix  $u_i = \overline{p}_i^\circ$  with  $\delta^{u_i u_i} = \delta^{u_i}$ . It follows that  $q_i \xrightarrow{p_i} r_i$  for  $0 \leq i \leq m'$ . Moreover, we have the following decompositions.

$$\begin{aligned} w_0 \overline{p}_0 \cdots w_{m+1} \overline{p}_{m+1} &= \underbrace{w_0}_{u_0} \underbrace{\overline{p}_0}_{v'_0 u_0} \underbrace{w_1}_{w'_0 u_1} \underbrace{\overline{p}_1}_{v''_0 v'_1 u_1} \underbrace{w_2}_{w'_1 u_2} \underbrace{\overline{p}_2}_{v''_1 v'_2 u_2} \cdots \underbrace{\overline{p}_m}_{v''_{m-1} v'_m u_m} \underbrace{w_{m+1}}_{w'_m u_{m+1}} \underbrace{\overline{p}_{m+1}}_{v''_m u_{m+1}} \\ &= \underbrace{u_0}_{u} \underbrace{v'_0 u_0 w'_0 u_1 v''_0 v'_1 u_1 w'_1 u_2 v''_1 v'_2 u_2 \cdots v''_{m-1} v'_m u_m w'_m u_{m+1} v''_m u_{m+1}}_{vv} \underbrace{u_{m+1}}_u \\ &= uv^{2(m+1)}u \end{aligned}$$

Since  $s_1$  has a  $u$ -loop and a  $v$ -loop it follows that  $r_{m'} = \delta(s_1, w_0 \overline{p}_0 \cdots w_{m+1} \overline{p}_{m+1}) = s_1$ . Hence we have shown that  $s_1 \xrightarrow{w_0} q_0 \xrightarrow{p_0} r_0 \xrightarrow{w_1} q_1 \xrightarrow{p_1} r_1 \xrightarrow{w_2} \cdots \xrightarrow{w_{m'}} q_{m'} \xrightarrow{p_{m'}} r_{m'} = s_1$ . By Definition 3.4 this means that  $p$  appears at  $s_1$ . Analogously we show that  $p$  appears at  $s_2$ .

Now define  $\tilde{q}_i =_{\text{def}} \delta(s_1, w_0 \cdots w_i)$  for  $0 \leq i \leq m'$ . Since  $s_1$  has a  $u$ -loop it holds that  $\tilde{q}_{m'} = \delta(s_1, w_0 \cdots w_{m'}) = \delta(s_1, uw'u) = \delta(s_1, w'u) = \delta(s_1, w) = s_2$ . This shows  $s_1 \xrightarrow{w_0} \tilde{q}_0 \xrightarrow{w_1} \tilde{q}_1 \xrightarrow{w_2} \cdots \xrightarrow{w_{m'}} \tilde{q}_{m'} = s_2$ . Moreover,  $p_i$  appears at  $\tilde{q}_i$  for  $0 \leq i \leq m'$  since  $w_i$  leads to a  $\overline{p}_i^\circ$ -loop in  $\mathcal{M}$ . Together with the fact that  $p$  appears at  $s_1$  and at  $s_2$  this implies  $s_1 \xrightarrow{p} s_2$  by Definition 3.4.  $\square$

**Lemma 4.49.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA. Suppose that there exist states  $s_1, s_2$  and words  $x, z \in A^*$ ,  $v, w \in A^+$  such that  $\alpha_0(w) \subseteq \alpha_0(v)$ ,  $s_0 \xrightarrow{x} s_1 \xrightarrow{v} s_1 \xrightarrow{z} +$  and  $s_1 \xrightarrow{w} s_2 \xrightarrow{v} s_2 \xrightarrow{z} -$ . Then  $\mathcal{M}$  has a pattern from  $\mathbb{P}_1^c$ .*

*Proof.* First of all we observe the following.

**Claim.** *We may assume that  $\mathbf{p}_1(w) = \mathbf{p}_1(v)$  and that  $w$  is a subword of  $v$ .*

If  $v$  and  $w$  start with different letters then we can use  $vwv$  instead of  $w$  since  $\alpha_0(vwv) \subseteq \alpha_0(v)$  and  $s_1 \xrightarrow{vwv} s_2$ . Now assume that  $w$  and  $v$  start with the same letter, but  $w$  is not a subword of  $v$ . The condition  $\alpha_0(w) \subseteq \alpha_0(v)$  means that all letters from  $w$  appear also in  $v$ . Hence,  $w$  is a subword of  $v^{|w|}$ . We can use  $v^{|w|}$  instead of  $v$  since  $\alpha_0(w) \subseteq \alpha_0(v) = \alpha_0(v^{|w|})$ ,  $s_1 \xrightarrow{v^{|w|}} s_1$  and  $s_2 \xrightarrow{v^{|w|}} s_2$ . This proves the claim.

Let  $n =_{\text{def}} |w| - 1$  and choose suitable letters  $a_0, \dots, a_n \in A$  such that  $w = a_0 \cdots a_n$ . By our claim there exist words  $v_0, \dots, v_n \in A^*$  such that  $v = a_0 v_0 a_1 v_1 \cdots a_n v_n$ . For  $0 \leq i \leq n$  let  $p_i =_{\text{def}} (\varepsilon, v_i)$  and note that  $p_i \in \mathcal{L}$  (see Definition 3.36). Moreover, let  $p =_{\text{def}} (a_0, p_0, \dots, a_n, p_n)$  and note that  $p \in \mathbb{P}_1^c$  (see Definition 3.4). Now we are going to show that  $s_1 \xrightarrow{p} s_2$ .

Note that by Definition 3.2,  $s_1 \xrightarrow{y} s_2$  implies  $s_1 \xrightarrow{(\varepsilon, y)} s_2$  for all  $y \in A^*$ . Therefore, if we define  $q_i =_{\text{def}} \delta(s_1, a_0 v_0 \cdots a_{i-1} v_{i-1} a_i)$  and  $r_i =_{\text{def}} \delta(q_i, v_i)$  for  $0 \leq i \leq n$  then we obtain

$$s_1 \xrightarrow{a_0} q_0 \xrightarrow{p_0} r_0 \xrightarrow{a_1} q_1 \xrightarrow{p_1} r_1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} q_n \xrightarrow{p_n} r_n = s_1.$$

Hence  $p$  appears at  $s_1$  by Definition 3.4. Analogously we show that  $p$  appears at  $s_2$ .

Let  $\tilde{q}_i =_{\text{def}} \delta(s_1, a_0 \cdots a_i)$  for  $0 \leq i \leq n$  and observe that  $s_1 \xrightarrow{a_0} \tilde{q}_0 \xrightarrow{a_1} \cdots \xrightarrow{a_n} \tilde{q}_n = s_2$ . It is easy to see that  $p_i = (\varepsilon, v_i)$  appears at each state in  $\mathcal{M}$ , and in particular at the states  $\tilde{q}_i$ . Together with the fact that  $p$  appears at  $s_1$  and at  $s_2$  this yields  $s_1 \xrightarrow{p} s_2$ . From Definition 3.5 it follows that  $\mathcal{M}$  has a pattern from  $\mathbb{P}_1^c$ .  $\square$

The main theorems of this subsection are given below. They establish the connection between the forbidden-pattern class  $\mathcal{FP}(\mathbb{P}_1^B)$  (respectively,  $\mathcal{FP}(\mathbb{P}_1^c)$ ) and the word extensions  $\preceq_{\mathcal{M}}^{1,k}$  (respectively,  $\preceq_{\mathcal{M}}^{1,0}$ ).

**Theorem 4.50.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA with  $L(\mathcal{M}) \in \mathcal{FP}(\mathbb{P}_1^B)$  and let  $k =_{\text{def}} 3 \cdot \mathcal{I}_{\mathcal{M}}$ . Then  $L(\mathcal{M})$  is a  $\preceq_{\mathcal{M}}^{1,k}$  co-ideal.*

*Proof.* Let  $n =_{\text{def}} \mathcal{I}_{\mathcal{M}}$  and assume that  $L(\mathcal{M})$  is not a  $\preceq_{\mathcal{M}}^{1,k}$  co-ideal, this will lead to a contradiction. Hence, there exist words  $y, y' \in A^+$ ,  $v \in W_{\mathcal{M}}^{1,k}$  such that  $y \in L(\mathcal{M})$ ,  $y' \notin L(\mathcal{M})$  and  $y <_v^{1,k} y'$ . This means that there exist words  $x, z \in A^*$  and  $w \in A^{*\geq k+1}$  such that  $y = xwz$ ,  $y' = xvwz$  and  $\alpha_k(vwv) \subseteq \alpha_k(v)$ . By the definition of  $W_{\mathcal{M}}^{1,k}$  there exist words  $r_1, \dots, r_n \in A^{*\geq k+1}$  such that  $v = r_1 r_2 \cdots r_n$  and  $\alpha_k(r_i) = \alpha_k(v)$  for  $1 \leq i \leq n$ . By Corollary 1.16, there exist  $i, j$  with  $1 \leq i \leq j \leq n$  such that  $\delta^{uu} = \delta^u$  for  $u =_{\text{def}} r_i r_{i+1} \cdots r_j$ .

Let  $x' =_{\text{def}} x r_1 \cdots r_{i-1} u$ ,  $z' =_{\text{def}} r_{j+1} \cdots r_n z$  and  $w' =_{\text{def}} r_{j+1} \cdots r_n w r_1 \cdots r_{i-1} u$ . Then  $y = x' z'$  and  $y' = x' w' z'$ . Since (i)  $r_i$  is a factor of  $u$  and (ii)  $u$  is a factor of  $v$ , it holds that  $\alpha_k(u) = \alpha_k(r_i) = \alpha_k(v)$ . Moreover, since  $uw'u$  is a factor of  $vwv$ , we have  $\alpha_k(uw'u) = \alpha_k(vwv) = \alpha_k(v) = \alpha_k(u)$ . So we have  $x', z' \in A^*$  and  $w', u \in A^{*\geq k+1}$  with

$\alpha_k(uw'u) = \alpha_k(u) \subseteq \alpha_k(uu)$ . Note that  $x'$  and  $w'$  lead to a  $u$ -loop in  $\mathcal{M}$  because both words have  $u$  as a suffix and  $\delta^{uu} = \delta^u$ . Therefore, with  $s_1 =_{\text{def}} \delta(s_0, x')$  and  $s_2 =_{\text{def}} \delta(s_1, w')$  we obtain  $s_0 \xrightarrow{x'} s_1 \xrightarrow{u} s_1 \xrightarrow{z'} +$  and  $s_1 \xrightarrow{w'} s_2 \xrightarrow{u} s_2 \xrightarrow{z'} -$ . From Lemma 4.48 it follows that  $\mathcal{M}$  has a pattern from  $\mathbb{P}_1^{\mathcal{B}}$ . By Definition 3.13 this is a contradiction to the assumption that  $L(\mathcal{M}) \in \mathcal{FP}(\mathbb{P}_1^{\mathcal{B}})$ .  $\square$

Theorem 4.50 can be easily transferred to the class  $\mathcal{FP}(\mathbb{P}_1^{\mathcal{L}})$ . For this we only have to use 0 instead of  $k$  in the proof above.

**Theorem 4.51.** *Let  $\mathcal{M} = (A, S, \delta, s_0, S')$  be a DFA with  $L(\mathcal{M}) \in \mathcal{FP}(\mathbb{P}_1^{\mathcal{L}})$ . Then  $L(\mathcal{M})$  is a  $\preceq_{\mathcal{M}}^{1,0}$  co-ideal.*

*Proof.* Let  $n =_{\text{def}} \mathcal{I}_{\mathcal{M}}$  and assume that  $L(\mathcal{M})$  is not a  $\preceq_{\mathcal{M}}^{1,0}$  co-ideal, this will lead to a contradiction. Hence, there exist words  $y, y' \in A^+$ ,  $v \in W_{\mathcal{M}}^{1,0}$  such that  $y \in L(\mathcal{M})$ ,  $y' \notin L(\mathcal{M})$  and  $y <_v^{1,0} y'$ . This means that there exist words  $x, z \in A^*$  and  $w \in A^+$  such that  $y = xwz$ ,  $y' = xwvz$  and  $\alpha_0(vwv) \subseteq \alpha_0(v)$ . By the definition of  $W_{\mathcal{M}}^{1,0}$  there exist words  $r_1, \dots, r_n \in A^+$  such that  $v = r_1 r_2 \cdots r_n$  and  $\alpha_0(r_i) = \alpha_0(v)$  for  $1 \leq i \leq n$ . By Corollary 1.16, there exist  $i, j$  with  $1 \leq i \leq j \leq n$  such that  $\delta^{uu} = \delta^u$  for  $u =_{\text{def}} r_i r_{i+1} \cdots r_j$ .

Let  $x' =_{\text{def}} x r_1 \cdots r_{i-1} u$ ,  $z' =_{\text{def}} r_{j+1} \cdots r_n z$  and  $w' =_{\text{def}} r_{j+1} \cdots r_n w r_1 \cdots r_{i-1} u$ . We obtain  $y = x' z'$  and  $y' = x' w' z'$ . Since (i)  $r_i$  is a factor of  $u$  and (ii)  $u$  is a factor of  $v$ , it holds that  $\alpha_0(u) = \alpha_0(r_i) = \alpha_0(v)$ . Moreover, since  $w'$  is a factor of  $vwv$ , we have  $\alpha_0(w') \subseteq \alpha_0(vwv) \subseteq \alpha_0(v) = \alpha_0(u)$ . This shows that  $x', z' \in A^*$  and  $w', u \in A^+$  with  $\alpha_0(w') \subseteq \alpha_0(u)$ . Note that  $x'$  and  $w'$  lead to a  $u$ -loop in  $\mathcal{M}$  because both words have  $u$  as a suffix and  $\delta^{uu} = \delta^u$ . Therefore, with  $s_1 =_{\text{def}} \delta(s_0, x')$  and  $s_2 =_{\text{def}} \delta(s_1, w')$  we obtain  $s_0 \xrightarrow{x'} s_1 \xrightarrow{u} s_1 \xrightarrow{z'} +$  and  $s_1 \xrightarrow{w'} s_2 \xrightarrow{u} s_2 \xrightarrow{z'} -$ . So we can apply Lemma 4.49 and it follows that  $\mathcal{M}$  has a pattern from  $\mathbb{P}_1^{\mathcal{L}}$ . By Definition 3.13 this is a contradiction to the assumption that  $L(\mathcal{M}) \in \mathcal{FP}(\mathbb{P}_1^{\mathcal{L}})$ .  $\square$

### 4.3.5 $\mathcal{L}_{3/2}$ and $\mathcal{B}_{3/2}$ are decidable

We combine the results of the preceding subsections and get the inclusion relations  $\mathcal{L}_{3/2} \supseteq \mathcal{FP}(\mathbb{P}_1^{\mathcal{L}})$  and  $\mathcal{B}_{3/2} \supseteq \mathcal{FP}(\mathbb{P}_1^{\mathcal{B}})$ . Together with the forbidden-pattern theory in chapter 3 this shows the forbidden-pattern characterizations  $\mathcal{L}_{3/2} = \mathcal{FP}(\mathbb{P}_1^{\mathcal{L}})$  and  $\mathcal{B}_{3/2} = \mathcal{FP}(\mathbb{P}_1^{\mathcal{B}})$ . Moreover, we obtain that the membership problems for  $\mathcal{L}_{3/2}$  and  $\mathcal{B}_{3/2}$  are decidable in nondeterministic logarithmic space.

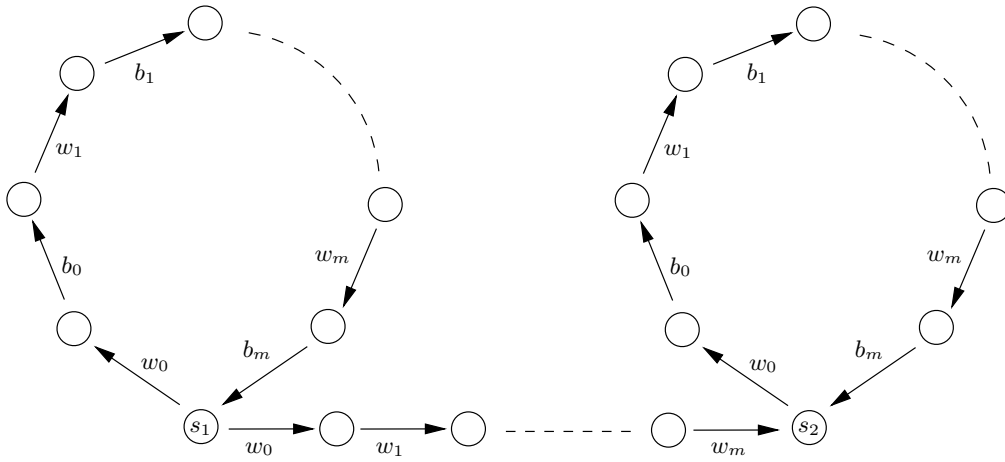
The decidability of  $\mathcal{L}_{3/2}$  and a forbidden-pattern characterization for this class was first given in [Arf91, PW97]. Note that the patterns given there and the patterns from  $\mathbb{P}_1^{\mathcal{L}}$  can be easily transformed into each other. Recently, it was shown in [PW01] that level  $n + 1/2$  of the DDH is decidable if and only if level  $n + 1/2$  of the STH is decidable (see Theorem 2.14). Together with [Arf91, PW97] this yields another proof for the decidability of  $\mathcal{B}_{3/2}$ .

**Theorem 4.52.** *It holds that  $\mathcal{L}_{3/2} = \mathcal{FP}(\mathbb{P}_1^{\mathcal{L}})$  and  $\mathcal{B}_{3/2} = \mathcal{FP}(\mathbb{P}_1^{\mathcal{B}})$ .*

*Proof.* By Theorem 3.39 it suffices to show  $\mathcal{FP}(\mathbb{P}_1^c) \subseteq \mathcal{L}_{3/2}$  and  $\mathcal{FP}(\mathbb{P}_1^B) \subseteq \mathcal{B}_{3/2}$ . Let  $L \in \mathcal{FP}(\mathbb{P}_1^c)$ ,  $L' \in \mathcal{FP}(\mathbb{P}_1^B)$  and let  $\mathcal{M}, \mathcal{M}'$  be DFAs such that  $L = L(\mathcal{M})$  and  $L' = L(\mathcal{M}')$ . By the Theorems 4.50 and 4.51, it holds that (i)  $L$  is a  $\preceq_{\mathcal{M}}^{1,0}$  co-ideal and (ii)  $L'$  is a  $\preceq_{\mathcal{M}'}^{1,k}$  co-ideal for  $k =_{\text{def}} 3 \cdot \mathcal{I}_{\mathcal{M}'}$ . From Theorem 4.46 it follows that there exist finite sets  $D, D' \subseteq A^+$  with

$$L = \langle D \rangle_{\preceq_{\mathcal{M}}^{1,0}} = \bigcup_{y \in D} \langle y \rangle_{\preceq_{\mathcal{M}}^{1,0}} \quad \text{and} \quad L' = \langle D' \rangle_{\preceq_{\mathcal{M}'}^{1,k}} = \bigcup_{y \in D'} \langle y \rangle_{\preceq_{\mathcal{M}'}^{1,k}}.$$

Finally, from Theorem 4.41 we get that  $\langle y \rangle_{\preceq_{\mathcal{M}}^{1,0}} \in \mathcal{L}_{3/2}$  and  $\langle y \rangle_{\preceq_{\mathcal{M}'}^{1,k}} \in \mathcal{B}_{3/2}$  for all  $y \in A^+$ . Since the unions above are finite, we obtain  $L \in \mathcal{L}_{3/2}$  and  $L' \in \mathcal{B}_{3/2}$ .  $\square$



**Fig. 4.5.** Forbidden-pattern for  $\mathcal{L}_{3/2}$  [PW97] with  $b_i \in A^*$  and  $w_i \in A^+$ .

**Corollary 4.53.** *For every DFA  $\mathcal{M} = (A, S, \delta, s_0, S')$  the following holds.*

$L(\mathcal{M}) \in \mathcal{L}_{3/2} \iff$  *there do not exist  $s_1, s_2 \in S$ ,  $z \in A^*$  such that  $s_0 \xrightarrow{z} s_1$ ,  $s_2 \xrightarrow{z} s_1$  and we find a pattern according to Figure 4.5 between  $s_1$  and  $s_2$ .*

$L(\mathcal{M}) \in \mathcal{B}_{3/2} \iff$  *there do not exist  $s_1, s_2 \in S$ ,  $z \in A^*$  such that  $s_0 \xrightarrow{z} s_1$ ,  $s_2 \xrightarrow{z} s_1$  and we find a pattern according to Figure 4.6 between  $s_1$  and  $s_2$ .*

*Proof.* This follows from Theorem 4.52 and the definition of the forbidden-pattern classes (see Definitions 3.1–3.5).  $\square$

**Theorem 4.54.** *On input of a DFA  $\mathcal{M}$ , the questions  $L(\mathcal{M}) \in \mathcal{L}_{3/2}$  and  $L(\mathcal{M}) \in \mathcal{B}_{3/2}$  are decidable in nondeterministic logarithmic space.*

*Proof.* This is an immediate consequence of the Theorems 3.46 and 4.52.  $\square$

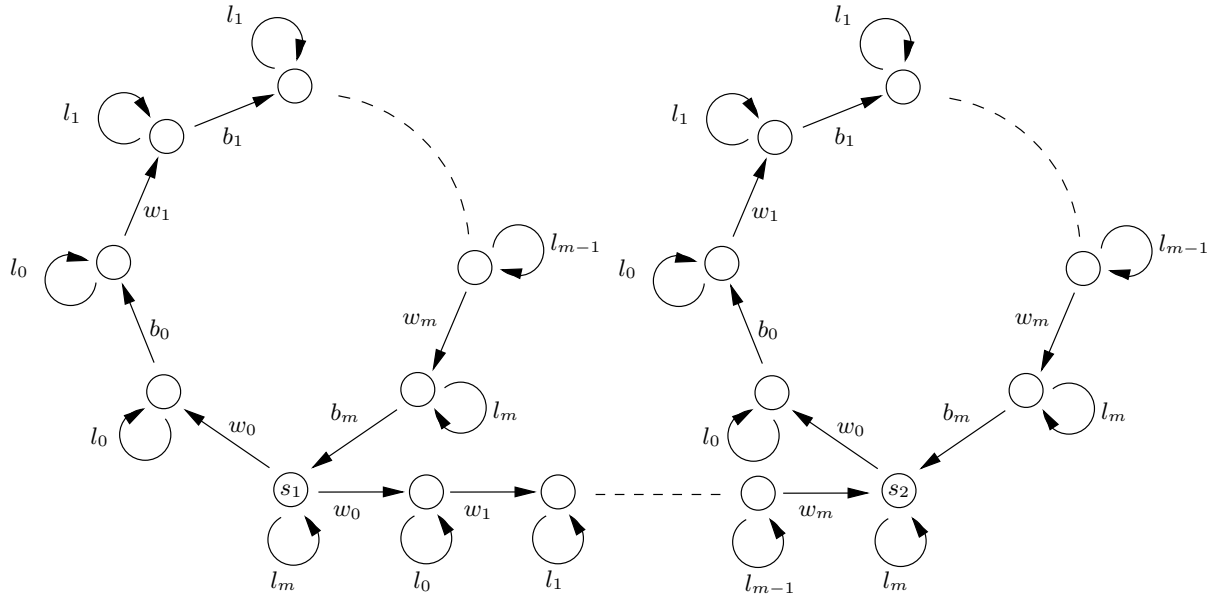


Fig. 4.6. Forbidden-pattern for  $\mathcal{B}_{3/2}$  with  $b_i, l_i, w_i \in A^+$ .

### 4.4 Summary and Discussion

In this chapter we showed that on the lower levels the classes of the concatenation hierarchies in fact coincide with the classes of the forbidden-pattern hierarchies from chapter 3. This refines our knowledge about these hierarchies and leads to Figure 4.7 which is an updated version of Figure 3.13. In particular this implies that the classes  $\mathcal{L}_{1/2}$ ,  $\mathcal{L}_{3/2}$ ,  $\mathcal{B}_{1/2}$  and  $\mathcal{B}_{3/2}$  are decidable in nondeterministic logarithmic space. Moreover, in this chapter we proved that the Boolean hierarchies over  $\mathcal{L}_{1/2}$  and over  $\mathcal{B}_{1/2}$  are decidable.

All these results were obtained by the use of certain word extensions. The advantage of this technique is that it makes possible to prove the decidability of the classes  $\mathcal{L}_{1/2}$ ,  $\mathcal{B}_{1/2}$ ,  $\mathcal{L}_{3/2}$  and  $\mathcal{B}_{3/2}$  in a uniform way. Moreover, this technique in combination with the method of alternating chains is ideally suited to attack decidability issues for the Boolean hierarchies over these classes. A very interesting starting point for future work is to clarify whether this approach is helpful when looking at the decidability of the Boolean hierarchy over  $\mathcal{B}_{3/2}$ .

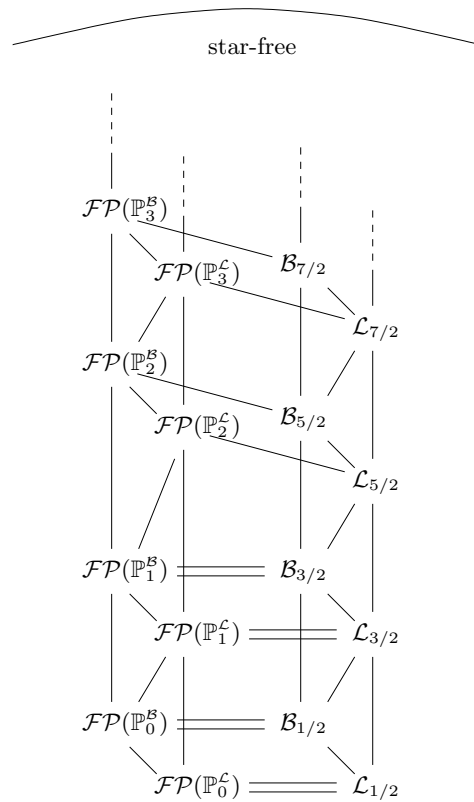


Fig. 4.7. Concatenation hierarchies and forbidden-pattern hierarchies



## Bibliography

- [Add65] J. Addison. The method of alternating chains. In *Theory of Models*, pages 1–16. North-Holland Publishing Co., 1965.
- [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.
- [BKS98] B. Borchert, D. Kuske, and F. Stephan. On existentially first-order definable languages and their relation to NP. In *Proc. 25rd International Colloquium on Automata, Languages and Programming*, volume 1443 of *Lecture Notes in Computer Science*, pages 17–28. Springer Verlag, 1998.
- [Brz76] J. A. Brzozowski. Hierarchies of aperiodic languages. *RAIRO Inform. Theor.*, 10:33–49, 1976.
- [Büc60] J. R. Büchi. Weak second-order arithmetic and finite automata. *Z. Math. Logik und Grundle. Math.*, 6:66–92, 1960.
- [CB71] R. S. Cohen and J. A. Brzozowski. Dot-depth of star-free events. *Journal of Computer and System Sciences*, 5:1–16, 1971.
- [CGH<sup>+</sup>88] J.-Y. Cai, T. Gundermann, J. Hartmanis, L. A. Hemachandra, V. Sewelson, K. W. Wagner, and G. Wechsung. The boolean hierarchy I: Structural properties. *SIAM Journal on Computing*, 17:1232–1252, 1988.
- [CH91] S. Cho and D. T. Huynh. Finite-automaton aperiodicity is PSPACE-complete. *Theoretical Computer Science*, 88:99–116, 1991.
- [CK96] C. Choffrut and J. Karhumäki. Combinatorics of words. In G. Rozenberg and A. Salomaa, editors, *Handbook of formal languages*, volume I, pages 329–438. Springer, 1996.
- [Eil76] S. Eilenberg. *Automata, languages and machines*, volume B. Academic Press, New York, 1976.
- [Ers68a] Y. L. Ershov. A hierarchy of sets I. *Algebra i Logika*, 7(1):47–74, 1968. In Russian.
- [Ers68b] Y. L. Ershov. A hierarchy of sets II. *Algebra i Logika*, 7(4):15–47, 1968. In Russian.

- [Gla98] C. Glaßer. A normalform for classes of concatenation hierarchies. Technical Report 216, Inst. für Informatik, Univ. Würzburg, 1998. Available via ftp from <http://www.informatik.uni-wuerzburg.de/reports/tr.html>.
- [Gla99] C. Glaßer. The boolean hierarchy over dot-depth 1/2. Technical Report 230, Inst. für Informatik, Univ. Würzburg, 1999. Available via ftp from <http://www.informatik.uni-wuerzburg.de/reports/tr.html>.
- [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>.
- [GS00a] C. Glaßer and H. Schmitz. Languages of dot-depth 3/2. In *Proceedings 17th Symposium on Theoretical Aspects of Computer Science*, volume 1770 of *Lecture Notes in Computer Science*, pages 555–566. Springer Verlag, 2000.
- [GS00b] C. Glaßer and H. Schmitz. Concatenation hierarchies and forbidden patterns. Technical Report 256, Inst. für Informatik, Univ. Würzburg, 2000.
- [GS00c] C. Glaßer and H. Schmitz. The boolean structure of dot-depth one. 2000. (To appear in *Journal of Automata, Languages and Combinatorics*).
- [GS00d] C. Glaßer and H. Schmitz. Decidable hierarchies of starfree languages. In *Proceedings 20th Conference on the Foundations of Software Technology and Theoretical Computer Science*, volume 1974 of *Lecture Notes in Computer Science*, pages 503–515. Springer Verlag, 2000.
- [GS01] C. Glaßer and H. Schmitz. Level 5/2 of the Straubing-Thérien hierarchy for two-letter alphabets. In *Preproceedings 5th Conference on Developments in Language Theory*, pages 254–265, 2001.
- [Hig52] G. Higman. Ordering by divisibility in abstract algebras. In *Proc. London Math. Soc.*, volume 3, pages 326–336, 1952.
- [Imm88] N. Immerman. Nondeterministic space is closed under complementation. *SIAM Journal on Computing*, 17:935–938, 1988.
- [Kle56] S. C. Kleene. Representation of events in nerve nets and finite automata. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, pages 3–42. Princeton University Press, Princeton, New Jersey, 1956.
- [Kna83] R. Knast. A semigroup characterization of dot-depth one languages. *RAIRO Inform. Théor.*, 17:321–330, 1983.
- [KSW87] J. Köbler, U. Schöning, and K. W. Wagner. The difference and the truth-table hierarchies for NP. *RAIRO Inform. Théor.*, 21:419–435, 1987.
- [MP71] R. McNaughton and S. Papert. *Counterfree Automata*. MIT Press, Cambridge, 1971.
- [Pap94] C. H. Papadimitriou. *Computational Complexity*. Addison-Wesley, Reading, MA, 1994.
- [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.
- [Pin96a] J. E. Pin. Logic, semigroups and automata on words. *Annals of Mathematics and Artificial Intelligence*, 16:343–384, 1996.

- [Pin96b] 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 *Lecture Notes in Computer Science*, 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.
- [PW01] J. E. Pin and P. Weil. The wreath product principle for ordered semigroups. 2001. To appear in *Communications in Algebra*. Available via ftp from <http://www.liafa.jussieu.fr/~jep>.
- [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.
- [Sch01] H. Schmitz. *The Forbidden-Pattern Approach to Concatenation Hierarchies*. PhD thesis, Fakultät für Mathematik und Informatik, Universität Würzburg, 2001.
- [Sel01] V. L. Selivanov. A logical approach to decidability of hierarchies of regular star-free languages. In *Proceedings 18th Symposium on Theoretical Aspects of Computer Science*, volume 2010 of *Lecture Notes in Computer Science*, pages 539–550. Springer Verlag, 2001.
- [Shu98] A. G. Shukin. Difference hierarchies of regular languages. *Theory of computing systems*, 161:141–155, 1998. In Russian.
- [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.
- [SS83] J. Sakarovitch and I. Simon. Subwords. In M. Lothaire, editor, *Combinatorics on Words*, Encyclopedia of mathematics and its applications, pages 105–142. Addison-Wesley, 1983.
- [SS00] V. L. Selivanov and A. G. Shukin. On hierarchies of regular star-free languages. Technical Report Preprint 69, A. P. Ershov Institute of Informatics Systems, Novosibirsk, 2000.
- [Ste85a] J. Stern. Characterizations of some classes of regular events. *Theoretical Computer Science*, 35:17–42, 1985.
- [Ste85b] J. Stern. Complexity of some problems from the theory of automata. *Information and Control*, 66:163–176, 1985.
- [Str81] H. Straubing. A generalization of the Schützenberger product of finite monoids. *Theoretical Computer Science*, 13:137–150, 1981.
- [Str85] H. Straubing. Finite semigroups 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-Thérien 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.
- [Tho96] W. Thomas. Languages, automata, and logic. In G. Rozenberg and A. Salomaa, editors, *Handbook of formal languages*, volume III, pages 389–456. Springer, 1996.
- [Tra61] B. A. Trakhtenbrot. Finite automata and logic of monadic predicates. *Doklady Akademii Nauk SSSR*, 140:326–329, 1961. In Russian.

## Notations

$\prec_v^{0,k}$	elementary word extensions which are used for the levels 1/2 of concatenation hierarchies (see Definition 1.18), <b>22</b> , 21–32, 78, 79, 88
$\prec_v^{1,k}$	elementary word extensions which are used for the levels 3/2 of concatenation hierarchies (see Definition 1.18), <b>22</b> , 21–32, 100, 103–105
$\succeq_{\mathcal{M}}^{0,k}$	word extensions which are used to obtain decidability results for the levels 1/2 of concatenation hierarchies (see Definition 4.2), <b>78</b> , 77–98
$\succeq_{\mathcal{M}}^{1,k}$	word extensions which are used to obtain decidability results for the levels 3/2 of concatenation hierarchies (see Definition 4.32), 77, <b>100</b> , 98–118
$\preceq$	the subword relation, $v \preceq w$ means that $v$ is a subword of $w$ , <b>13</b> , 83
$\preceq_{\mathcal{M}}^{0,k}$	this weakened version of the word extension $\succeq_{\mathcal{M}}^{0,k}$ is used in the proof of Theorem 4.8, <b>83–85</b>
$\triangleleft$	a relation for pattern classes, $\mathbb{P}_{n_1}^{X_1} \triangleleft \mathbb{P}_{n_2}^{X_2}$ means that every pattern from $\mathbb{P}_{n_2}^{X_2}$ can be interpreted as a pattern from $\mathbb{P}_{n_1}^{X_1}$ , <b>54</b> , 68
$\triangleleft_k$	the $k$ -embedding from [Ste85a] (see also [Sch01, section 2.7] and [Sch01, Definition 2.1] for a discussion and for an equivalent definition), this notation is only used in the proof of Proposition 4.29, 98
$\sim_{\mathcal{M}}$	$v \sim_{\mathcal{M}} w$ means that $\delta^v = \delta^w$ where $\delta$ is the transition function of $\mathcal{M}$ , <b>14</b>
$\langle s \rangle_{<}$	the $<$ upward closure of an element $s$ for an arbitrary binary relation $<$ , <b>16</b>
$\langle T \rangle_{<}$	the $<$ upward closure of a set $T$ for an arbitrary binary relation $<$ , <b>16</b>
$\langle y \rangle_{\prec_v^{1,k}}$	the $\prec_v^{1,k}$ upward closure of a word $y$ , 25, 27–29, 32, 106, 107
$\langle y \rangle_{\succeq_{\mathcal{M}}^{0,k}}$	the $\succeq_{\mathcal{M}}^{0,k}$ upward closure of a word $y$ , 80–83
$\langle y \rangle_{\succeq_{\mathcal{M}}^{1,k}}$	the $\succeq_{\mathcal{M}}^{1,k}$ upward closure of a word $y$ , 107, 108, 113
$(\beta \Gamma \delta)_k$	is defined as $\{ w \in A^{*\geq k+1} \mid \mathfrak{p}_k(w) = \beta, \mathfrak{s}_k(w) = \delta, \alpha_k(w) \subseteq \Gamma \}$ for $k \geq 0$ , $\beta, \delta \in A^k$ and $\Gamma \subseteq A^{k+1}$ , <b>23</b>
$\lfloor r \rfloor$	is the greatest integer that is less than or equal to $r$ , <b>15</b>
$w^\infty$	denotes the infinite word $a_1 \cdots a_m a_1 \cdots a_m \cdots$ where the $a_i$ are alphabet letters with $w = a_1 \cdots a_m$ , <b>56</b> , 58
$ w $	the length of the word $w$ , <b>13</b>
$w[i, j]$	choose suitable letters $a_i \in A$ such that $w = a_1 \cdots a_n$ , then $w[i, j] =_{\text{def}} a_i a_{i+1} \cdots a_{j-1}$ for $1 \leq i \leq j \leq  w  + 1$ , <b>14</b>
$Lw^{-1}$	the right residual of $L$ , <b>14</b>
$w^{-1}L$	the left residual of $L$ , <b>13</b>
$A^{-j}w$	the word that emerges from $w$ when deleting the first $j$ letters, <b>14</b>
$wA^{-j}$	the word that emerges from $w$ when deleting the last $j$ letters, <b>14</b>
$\bar{p}$	the bridge-word of a pattern $p$ (see Definition 3.6), <b>46</b> , 58

$\bar{p}^\circ$	the loop-word of a pattern $p$ (see Definition 3.6), <b>46</b>
$ \mathcal{M} $	denotes the size of the DFA $\mathcal{M}$ , i.e., the number of states, <b>14</b>
$s_1 \xrightarrow{z} s_2$	means that $\delta(s_1, z) = s_2$ for a fixed DFA $\mathcal{M}$ with transition function $\delta$ , <b>14</b>
$s_1 \longrightarrow s_2$	means that there exists a $z \in A^*$ with $s_1 \xrightarrow{z} s_2$ , <b>14</b>
$s_1 \xrightarrow{z} +$	means that there exists an accepting state $s_2$ with $s_1 \xrightarrow{z} s_2$ , <b>14</b>
$s_1 \xrightarrow{z} -$	means that there exists a rejecting state $s_2$ with $s_1 \xrightarrow{z} s_2$ , <b>14</b>
$s_1 \xleftrightarrow[a]{b} s_2$	abbreviation for $s_1 \xrightarrow{a} s_2$ and $s_2 \xrightarrow{b} s_1$ for letters $a$ and $b$ , <b>71</b>
$s_1 \overset{p}{\leftrightarrow} s_2$	means that the states $s_1$ and $s_2$ are connected via pattern $p$ (see Definitions 3.2 and 3.4), <b>45, 47, 60, 71</b>
$A$	finite alphabet with at least two letters, <b>13</b>
$A^*$	set of all words over $A$ , <b>13</b>
$A^+$	set of nonempty words over $A$ , <b>13</b>
$A^k$	set of words with length = $k$ , <b>13</b>
$A^{\geq k}$	set of words with length $\geq k$ , <b>13</b>
$A^{\leq k}$	set of words with length $\leq k$ , <b>13</b>
$A^{+\geq k}$	set of nonempty words with length $\geq k$ , <b>13</b>
$A^{+\leq k}$	set of nonempty words with length $\leq k$ , <b>13</b>
$\alpha_k(w)$	set of factors of length $k + 1$ in $w$ , <b>13, 107, 114–116</b>
$\mathcal{B}$	denotes the class of initial patterns that corresponds to level 1/2 of the dot-depth hierarchy (see Definition 3.36), <b>68, 85</b>
$\mathbb{B}_{1/2}$	denotes the forbidden-pattern for level 1/2 of the dot-depth hierarchy (see Figure 3.1), <b>42</b>
$\mathbb{B}_{3/2}$	denotes the forbidden-pattern for level 3/2 of the dot-depth hierarchy (see Figure 3.2), <b>42</b>
$\text{BC}(\mathcal{C})$	the Boolean closure of a class of languages $\mathcal{C}$ , i.e., the closure of $\mathcal{C}$ under union, intersection and complementation, <b>8, 15, 34</b>
$\text{BC}^*(\mathcal{D})$	alternative version of the Boolean closure of a class of languages $\mathcal{D}$ (here $A^*$ is considered as the universe, i.e., complementation is taken w.r.t. $A^*$ ), <b>36</b>
$\tilde{\mathcal{B}}_k$	certain classes of languages (see Definition 1.21) which are contained in level 3/2 of the dot-depth hierarchy, <b>23, 22–32</b>
$\tilde{\mathcal{B}}_{k,m}$	classes of languages that refine the classes $\tilde{\mathcal{B}}_k$ (see Definition 1.24), <b>23</b>
$\mathcal{B}_{n/2}$	the classes (levels) of the dot-depth hierarchy (DDH), <b>8, 34, 35, 69, 71, 73</b>
$\mathcal{B}_{n/2}^+$	denotes the alternative classes (levels) of the dot-depth hierarchy from the literature, <b>36</b>
$\mathcal{C}_{\mathcal{M}}^k$	this function is used to bound the length of certain words in section 4.3 (see Definition 4.30), <b>99, 111, 112</b>
$\mathcal{C}(l)$	class of the Boolean hierarchy over $\mathcal{C}$ , consists of languages $L$ that can be written as $L = L_1 \setminus (L_2 \setminus (\dots \setminus L_l))$ for $L_1, L_2, \dots, L_l \in \mathcal{C}$ with $L_1 \supseteq L_2 \supseteq \dots \supseteq L_l$ , <b>15, 87, 89</b>
$\text{co}\mathcal{C}$	set of complements with respect to $A^+$ of the class of languages $\mathcal{C}$ , <b>13</b>
$\text{co}^*\mathcal{D}$	set of complements with respect to $A^*$ of the class of languages $\mathcal{C}$ , <b>36</b>
DDH	abbreviation for dot-depth hierarchy, <b>33</b>
$\delta^w$	this function is defined as $\delta^w(s) \stackrel{\text{def}}{=} \delta(s, w)$ for a word $w$ and a transition function $\delta$ , <b>14, 17–21</b>
DFA	abbreviation for deterministic finite automaton, <b>14</b>

$\mathcal{D}_{\mathcal{M}}^k(n)$	this function is used to bound the length of certain words in section 4.2 (see Definition 4.22), <b>95</b> , <b>96</b>
$\mathcal{E}_{\mathcal{M}}^k(n)$	this function is used to bound the length of certain words in section 4.2 (see Definition 4.22), <b>95</b> , <b>97</b>
$\varepsilon$	the empty word, <b>13</b>
$\mathcal{F}_{\mathcal{M}}^k(m)$	this function is used to bound the length of certain words in section 4.3 (see Definition 4.30), <b>99</b> , <b>111</b>
$\text{FO}[\lt]$	first-order logic with relation $\lt$ , <b>9</b>
$\text{FO}[\lt, \min, \max, S, P]$	first-order logic with relation $\lt$ , constants $\min, \max$ and functions $S$ (successor), $P$ (predecessor), <b>9</b>
$\mathcal{FP}(\mathbb{P}_n^{\mathcal{E}})$	class of languages that can be accepted by a DFA which does not have a pattern from $\mathbb{P}_n^{\mathcal{E}}$ , this class contains the level $n + 1/2$ of the dot-depth hierarchy, <b>67–75</b> , <b>85–87</b> , <b>114–118</b>
$\mathcal{FP}(\mathbb{P}_n^{\mathcal{I}})$	class of languages that can be accepted by a DFA which does not have a pattern from $\mathbb{P}_n^{\mathcal{I}}$ , <b>51</b> , <b>51–62</b>
$\mathcal{FP}(\mathbb{P}_n^{\mathcal{L}})$	class of languages that can be accepted by a DFA which does not have a pattern from $\mathbb{P}_n^{\mathcal{L}}$ , this class contains the level $n + 1/2$ of the Straubing-Thérien hierarchy, <b>67–75</b> , <b>85–87</b> , <b>114–118</b>
$\text{FU}(\mathcal{C})$	the closure of the class of languages $\mathcal{C}$ under finite (possibly empty) union, <b>34</b>
$\mathcal{I}$	denotes a class of initial patterns (see Definition 3.1), <b>44</b>
$\mathcal{I}_{\mathcal{M}}$	is defined as $( \mathcal{M}  + 1)^{( \mathcal{M} +1)^{( \mathcal{M} +1)}}$ where $\mathcal{M}$ is a DFA, <b>17</b> , <b>17–21</b> , <b>99</b>
$\text{IT}(\mathbb{P})$	denotes an iteration rule for patterns which is defined as $\text{IT}(\mathbb{P}) \stackrel{\text{def}}{=} \{(w_0, p_0, \dots, w_m, p_m) \mid m \geq 0, p_i \in \mathbb{P}, w_i \in A^+\}$ (see Definition 3.3), <b>45</b>
$\text{L}$	class of languages that can be accepted by a Turing machine in deterministic logarithmic space, <b>14</b>
$\mathcal{L}$	denotes the class of initial patterns that corresponds to level $1/2$ of the Straubing-Thérien hierarchy (see Definition 3.36), <b>68</b> , <b>85</b>
$\mathbb{L}_{1/2}$	denotes the forbidden-pattern for level $1/2$ of the Straubing-Thérien hierarchy (see Figure 3.3), <b>43</b>
$\mathbb{L}_{3/2}$	denotes the forbidden-pattern for level $3/2$ of the Straubing-Thérien hierarchy (see Figure 3.4), <b>43</b>
$\lambda(p)$	a transformation of the pattern $p$ (see Definition 3.8), <b>46</b> , <b>72</b>
$L_{k,v}$	is equal to $\{w \in A^{*\geq k+1} \mid \alpha_k(vwv) = \alpha_k(v)\}$ for $k \geq 0$ and $v \in A^{*\geq k+1}$ , <b>23</b>
$L(\mathcal{M})$	the language accepted by the DFA $\mathcal{M}$ , <b>14</b>
$\mathcal{L}_{n/2}$	the classes (levels) of the Straubing-Thérien hierarchy (STH), <b>8</b> , <b>34</b> , <b>35</b> , <b>69</b>
$\mathcal{L}_{n/2}^*$	denotes the alternative classes (levels) of the Straubing-Thérien hierarchy from the literature, <b>36</b>
$\mathcal{M}$	denotes a deterministic finite automaton, <b>14</b>
$n_{\mathcal{M}}^{0,k}$	denotes the maximal number of alternations (with respect to $L(\mathcal{M})$ ) in $\preceq_{\mathcal{M}}^{0,k}$ chains (see Definition 4.15), <b>87</b> , <b>88</b> , <b>89</b>
$m_{\mathcal{B}}(\mathcal{M})$	the minimal number $n$ such that the languages $L(\mathcal{M})$ belongs to the $n$ -th level of the Boolean hierarchy over $\mathcal{B}_{1/2}$ , <b>97</b> , <b>98</b>
$m_{\mathcal{L}}(\mathcal{M})$	the minimal number $n$ such that the languages $L(\mathcal{M})$ belongs to the $n$ -th level of the Boolean hierarchy over $\mathcal{L}_{1/2}$ , <b>97</b> , <b>98</b>
$(m \bmod n)$	abbreviation for $m - n\lfloor m/n \rfloor$ , <b>56</b>
$\text{NL}$	class of languages that can be accepted by a Turing machine in nondeterministic logarithmic space, <b>14</b> , <b>62–67</b>

NP	class of languages that can be accepted by a Turing machine in nondeterministic polynomial time, <b>14</b>
P	class of languages that can be accepted by a Turing machine in deterministic polynomial time, <b>14</b>
$\text{PATTERN}_{n,k}^{\mathcal{I}}$	this problem addresses the existence of patterns that appear simultaneously in a given DFA (see Definition 3.30), <b>63</b> , 62–67
$\mathcal{P}(B)$	the power set of a set $B$ , <b>15</b>
$\mathbb{P}_n^{\mathcal{B}}$	the class of patterns that is obtained from the class of initial patterns $\mathcal{B}$ by applying the iteration rule IT $n$ times (see Definition 3.4), this pattern class is used to define classes of languages containing the level $n+1/2$ of the dot-depth hierarchy, 67–75, 85, 114
$\mathbb{P}_n^{\mathcal{I}}$	the class of patterns that is obtained from the class of initial patterns $\mathcal{I}$ by applying the iteration rule IT $n$ times (see Definition 3.4), <b>45</b> , 51–61
$\mathbb{P}_n^{\mathcal{L}}$	the class of patterns that is obtained from the class of initial patterns $\mathcal{L}$ by applying the iteration rule IT $n$ times (see Definition 3.4), this pattern class is used to define classes of languages containing the level $n+1/2$ of the Straubing-Thérien hierarchy, 67–75, 85
$\pi(p, r)$	a transformation of the pattern $p$ (see Definition 3.10), <b>47</b> , 49
$\mathfrak{p}_k(w)$	the $k$ -prefix of the word $w$ , <b>14</b>
$\text{Pol}^{\mathcal{B}}(\mathcal{C})$	alternative version of the polynomial closure of a class of languages $\mathcal{C}$ , this version is used in the literature to define the dot-depth hierarchy, <b>36</b>
$\text{Pol}(\mathcal{C})$	polynomial closure of a class of languages $\mathcal{C}$ , 8, <b>34</b> , 52
$\text{Pol}^{\mathcal{L}}(\mathcal{C})$	alternative version of the polynomial closure of a class of languages $\mathcal{C}$ , this version is used in the literature to define the Straubing-Thérien hierarchy, <b>36</b>
PSPACE	class of languages that can be accepted by a Turing machine in deterministic polynomial space, <b>14</b>
$\text{REACH}_k$	this problem addresses the existence of paths that appear simultaneously in a given DFA (see Definition 3.29), <b>63</b> , 62–67
SF	the class of starfree languages, <b>33</b> , 35, 56–62, 70, 74
$\mathfrak{s}_k(w)$	the $k$ -suffix of the word $w$ , <b>14</b>
STH	abbreviation for Straubing-Thérien hierarchy, <b>33</b>
$v(i, j)$	this notation is only used in section 1.3 and is defined as $v_i v_{i+1} \cdots v_{j-1}$ for a given list of words $v_0, v_1, \dots, v_l \in A^+$ , <b>18</b>
$\mathbb{W}_{\mathcal{M}}^{0,k}$	set of context words for $\preceq_{\mathcal{M}}^{0,k}$ word extensions, <b>78</b> , 92, 94
$\mathbb{W}_{\mathcal{M}}^{1,k}$	set of context words for $\preceq_{\mathcal{M}}^{1,k}$ word extensions, <b>99</b> , 101, 109–113
wpos	abbreviation for well partial ordered set, <b>16</b> , 17, 113



# Index

- acceptance by a DFA, *see* DFA
- acceptance by a finite monoid, *see* finite monoid
- alternating chain, 119
- aperiodic syntactic monoid, 7
- appearance of a pattern, *see* pattern appears
  
- binary relation, 16
  - on the set of words, 16, 21
- Boolean closure, 15, **34**
  - alternative definitions, 36
- Boolean combination, 15
- Boolean hierarchy, **15**
  - over  $\mathcal{B}_{1/2}$ , 10, 77, 87–98
  - characterization, 88, 89, 95
  - computation of the exact level, 92, 97–98
  - decidability results, *see* decidability
  - strictness, 87, 91
  - over  $\mathcal{B}_{n+1/2}$ , strictness, 87, 91
  - over  $\mathcal{L}_{1/2}$ , 10, 77, 87–98
  - characterization, 88, 89, 95
  - computation of the exact level, 92, 97–98
  - decidability results, *see* decidability
  - strictness, 87, 91
  - over  $\mathcal{L}_{n+1/2}$ , strictness, 87, 91
- Boolean operation, 33
- bounding function
  - $\mathcal{C}_{\mathcal{M}}^k$ , **99**
  - $\mathcal{D}_{\mathcal{M}}^k(n)$ , **95**
  - $\mathcal{E}_{\mathcal{M}}^k(n)$ , **95**
  - $\mathcal{F}_{\mathcal{M}}^k(m)$ , **99**
  - $\mathcal{I}_{\mathcal{M}}$ , **17**, 17–21, 99
- bridge-word, **46**, 49, 58
  
- chain of word extensions, *see* word extension
- class of initial patterns, 41, **44**, 45
  - $\mathcal{B}$ , **68**
  - $\mathcal{L}$ , **68**
- class of patterns, *see* pattern class
- co-ideal, **16**, *see also* upward closure
  - finitely generated, **16**, 17, 85, 109, 113
  - generated by a set, **16**
  - generated by an element, **16**
  - with respect to  $\sum_{\mathcal{M}}^{0,k}$ , 83–85
  - with respect to  $\sum_{\mathcal{M}}^{1,k}$ , 109, 113–117
- concatenation, 33
- concatenation complexity, 33
  
- concatenation hierarchy, 8, 33, 119
  - inclusion structure, 120
- connected, *see* states are connected
- context word, 21, 78, 99
  - how to find, 109–113
  - rearrangement, 103–107
  - set of context words  $W_{\mathcal{M}}^{0,k}$ , **78**
  - set of context words  $W_{\mathcal{M}}^{1,k}$ , **99**
  
- decidability
  - $\mathcal{B}_{1/2}$ , 86, 87
  - $\mathcal{B}_{3/2}$ , 117, 118
  - Boolean hierarchy over  $\mathcal{B}_{1/2}$ , 97
  - Boolean hierarchy over  $\mathcal{L}_{1/2}$ , 97
  - forbidden-pattern classes  $\mathcal{FP}(\mathbb{P}_n^x)$ , 62–67
  - $\mathcal{FP}(\mathbb{P}_n^B)$ , 74
  - $\mathcal{FP}(\mathbb{P}_n^L)$ , 74
  - $\mathcal{L}_{1/2}$ , 86, 87
  - $\mathcal{L}_{3/2}$ , 117, 118
  - partial result for  $\mathcal{L}_2$ , 9
  - partial result for  $\mathcal{L}_{5/2}$ , 9
  - SF, 74
- decomposition of a word, 15
  - into idempotents, 17–21
- deterministic finite automaton, *see* DFA
- deterministic logarithmic space L, **14**, 63
- deterministic polynomial space PSPACE, **14**
  - PSPACE-complete, 74
- deterministic polynomial time P, **14**
- DFA, deterministic finite automaton, 7, **14**
  - accepted language, **14**
  - as input to a Turing machine, 63
  - has a nontrivial permutation, 9, 15
  - has a pattern, **45**, 51, 67, 72
    - from  $\mathbb{P}_1^B$ , 115
    - from  $\mathbb{P}_1^L$ , 116
  - initial state, **14**
  - input alphabet, **14**
  - minimal, **14**, 15
  - permutationfree, 7, 9, **15**
  - set of accepting states, **14**
  - set of states, **14**
  - transition function, **14**
- dot-depth, 8
- dot-depth hierarchy, 8, 33, **34**
  - alternative definition, **36**

- closure under intersection, 37
- closure under taking residuals, 37
- connection to Straubing-Thérien hierarchy, 35, 37, 120
- decidability results, *see* decidability
- exhaust the class of starfree languages, 35
- inclusion relations with respect to forbidden-pattern hierarchies, 69
- inclusion structure, 34, 35, 75
- lower bound algorithm, 42, 75
- normalforms, 37
- separation of  $\mathcal{B}_n$  and  $\mathcal{FP}(\mathbb{P}_n^c)$ , 73
- strictness, 35, 71
- upper bound for leaf-language from  $\mathcal{B}_1$ , *see* polynomial time leaf-language
- dot-depth problem, 8, **37**, 75, 77
  
- empty word  $\varepsilon$ , **13**
  
- factor of a word, **13**
- finite alphabet  $A$ , **13**
- finite automaton, *see* DFA
- finite monoid, acceptance by, 7
- finitely generated co-ideal, *see* co-ideal
- first-order logic, 8
  - FO[<], 9
  - FO[<, min, max,  $S$ ,  $P$ ], 9
- forbidden-pattern characterization, 9, 38, **41**
  - for  $\mathcal{B}_{1/2}$ , 39, 86
  - for  $\mathcal{B}_{3/2}$ , 40, 117, 119
  - for  $\mathcal{L}_{1/2}$ , 38, 86
  - for  $\mathcal{L}_{3/2}$ , 39, 117, 118
- forbidden-pattern class, 10, **51**
  - $\mathcal{FP}(\mathbb{P}_0^S)$ , 85
  - $\mathcal{FP}(\mathbb{P}_1^S)$ , 114, 116
  - $\mathcal{FP}(\mathbb{P}_0^c)$ , 85
  - $\mathcal{FP}(\mathbb{P}_1^c)$ , 114, 116, 117
- decidability results, *see* decidability
- forbidden-pattern hierarchy
  - decidability results, *see* decidability
  - exhaust the class of starfree languages, 61, 70
  - inclusion relations with respect to concatenation hierarchies, 69
  - inclusion structure, 54–56, 68, 75, 120
  - separation of  $\mathcal{B}_n$  and  $\mathcal{FP}(\mathbb{P}_n^c)$ , 73
  - strictness, 71–74
- forbidden-pattern theory, 41–76
- full levels, **34**
  
- graph accessibility problem, 64
  
- half levels, **34**
- hierarchy of pattern classes, **45**
  
- idempotent, 17–21
  - decomposition of a word into idempotents, 17–21
  - for a DFA, **14**
- infinite set of pairwise incomparable words, **16**
- infinite word, *see* periodic, infinite word
  
- infinite, strictly descending chain, **16**
- initial pattern, *see* class of initial patterns
- input tape, 64
- iteration, 33
- iteration rule IT, **45**, 41–62
  - versus polynomial closure, 52
  
- $k$ -prefix of a word, **14**, 79, 100
- $k$ -suffix of a word, **14**, 79, 100
  
- L, *see* deterministic logarithmic space
- language accepted by a DFA, *see* DFA
- leaf-language, *see* polynomial time leaf-language
- left residual, **13**
- length of a word, **13**
- loop-structure, **42**, 46
- loop-word, **46**, 49
- lower bound algorithm for the dot-depth, *see* dot-depth hierarchy
  
- mapping leads to, **14**, 18
- minimal DFA, *see* DFA
- monadic second-order logic, 7
  
- nondeterministic logarithmic space NL, **14**, 64, 67, 74, 87, 118
- nondeterministic polynomial time NP, **14**
  
- oracle machine, 64
- ordered set, **16**
  
- P, *see* deterministic polynomial time
- partial ordered set, **16**
- pattern appears at a state, **45**, 46, 47
  - example, 65
- pattern class, 41, **45**
- pattern connects two states, *see* states are connected
- pattern iteration, *see* iteration rule IT
- periodic, infinite word, **56**
- permutationfree, *see* DFA
- polynomial closure, 8, **34**, 46, 51, 52
  - alternative definitions, 36
  - versus pattern iteration, 52
- polynomial time leaf-language, 87
  - upper bound for leaf-language from  $\mathcal{B}_1$ , 87
- power set, **15**
- predecessor
  - with respect to  $\preceq_{\mathcal{M}}^{0,k}$ , 95
  - with respect to  $\preceq_{\mathcal{M}}^{1,k}$ , 109
- prefix of a word, *see*  $k$ -prefix of a word
- PSPACE, *see* deterministic polynomial space
  
- query tape, 64
  
- rearrangement of  $\prec_v^{1,k}$  chains, *see* context word
- rearrangement of context words, *see* context word
- regular expression, 7
- regular language, 7, **14**, 33
- right residual, **14**

- second-order logic, *see* monadic second-order logic
- separability of  $\mathcal{L}_{3/2}$ , 87
- set of all words, **13**
- set of context words, *see* context word
- set of nonempty words, **13**
- starfree languages, 8, 15, **33**, 35, 56, 70
  - decidability results, *see* decidability
- starfree regular expression, 7
- state has a loop, **14**
- states are connected via a pattern, **45**, 46, 47
  - example, 65
- Straubing-Thérien hierarchy, 8, 33, **34**
  - alternative definition, **36**
  - closure under intersection, 37
  - closure under taking residuals, 37
  - connection to dot-depth hierarchy, 35, 37, 120
  - decidability results, *see* decidability
  - exhaust the class of starfree languages, 35
  - inclusion relations with respect to forbidden-pattern hierarchies, 69
  - inclusion structure, 34, 35, 75
  - normalforms, 37
  - separability of  $\mathcal{L}_{3/2}$ , 87
  - strictness, 35
- subword relation  $\preceq$ , **13**, 83
- suffix of a word, *see*  $k$ -suffix of a word
- swap successive closure operations, *see* context word, rearrangement
- syntactic monoid, *see* aperiodic syntactic monoid
  
- technique of word extensions, *see* word extension
- transformation of  $\prec_v^{1,k}$  chains, *see* context word, rearrangement
- transformation of patterns, 46
- transition function, *see* DFA
- transposition, 105
  
- upward closure, **16**, *see also* co-ideal
  - with respect to  $\prec_v^{1,k}$ , 21–32, 106, 107
  - of a nonempty word, 24–27
  - of languages from  $\tilde{\mathcal{B}}_k$ , 29–32
  - of languages from  $\tilde{\mathcal{B}}_{k,1}$ , 28–29
  - with respect to  $\preceq_{\mathcal{M}}^{0,k}$ , 80–83
  - with respect to  $\preceq_{\mathcal{M}}^{1,k}$ , 103–109
  
- well partial ordered set, **16**, 83–85, 113
- word extension, 10, 21, 77, 119
  - $\prec_v^{0,k}$ , **22**, 21–32, 78–79, 88
  - $\prec_v^{1,k}$ , **22**, 21–32, 100–106
  - $\preceq_{\mathcal{M}}^{0,k}$ , **78**, 77–98
  - $\preceq_{\mathcal{M}}^{1,k}$ , **77**, **100**, 98–118
  - chain of  $\preceq_{\mathcal{M}}^{0,k}$  extensions, 87–88, 95, 97
  - bounding the length of words in a chain, 97
  - increasing the parameter  $k$ , 89
  - maximal number of alternations, **87**, 89, 95, 97
- word leads to, **14**, 49
- working tape, 64
- wpos, *see* well partial ordered set

