For submission to the Journal of Automata, Languages and Combinatorics
Created on March 5, 2018

POSITION AUTOMATA FOR SEMI-EXTENDED EXPRESSIONS

Sabine Broda António Machiavelo Nelma Moreira Rogério Reis

CMUP & DCC & DM
Faculdade de Ciências da Universidade do Porto,
Rua do Campo Alegre 1021, 4219-007, Portugal
sbb@dcc.fc.up.pt,ajmachia@fc.up.pt,{nam,rvr}@dcc.fc.up.pt

ABSTRACT

Positions and derivatives are two essential notions in the conversion methods from regular expressions to equivalent finite automata. Partial derivative based methods have recently been extended to regular expressions with intersection (semi-extended). In this paper, we present a position automaton construction for those expressions. This construction generalizes the notion of position, making it compatible with intersection. The resulting automaton is homogeneous and has the partial derivative automaton as a quotient.

1. Introduction

The position automaton ($\mathcal{A}_{\mathsf{pos}}$), introduced by Glushkov [14], permits the conversion of a simple regular expression (involving only the union, concatenation and star operations) into an equivalent nondeterministic finite automaton (NFA) without ε-transitions. The states in the position automaton correspond to the positions of letters in the corresponding regular expression plus an additional initial state. McNaughton and Yamada [17] also used the positions of a regular expression to define an automaton, however they directly computed a deterministic version of the position automaton. The position automaton has been well studied [4, 10] and it is considered the *standard* automaton simulation of a regular expression [18]. Some of its interesting properties are: homogeneity, i.e. for each state, all in-transitions have the same label (letter); whenever deterministic, these automata characterize certain families of unambiguous regular expressions; and can be computed in quadratic time [6]; other automata simulations of regular expressions are quotients of the $\mathcal{A}_{\mathsf{pos}}$, e.g. the partial derivative automata ($\mathcal{A}_{\mathsf{pd}}$) [11] and the follow automata [16].

Many authors observed that the position automaton construction could not directly be extended to regular expressions with intersection [4, 8], as intersection (and

This work was partially supported by CMUP (UID/MAT/00144/2013), which is funded by FCT (Portugal) with national (MEC) and European structural funds through the programs FEDER, under the partnership agreement PT2020.

also complementation) is not compatible with the notion of position. In fact, considering the positions of letters in the expression $(ab^\star) \cap a$, whose language is $\{a\}$, we obtain the regular expression $(a_1 b_2^\star) \cap a_3$. Interpreting a_1 and a_3 as distinct alphabet symbols, the language described by this expression is empty and there is no longer a correspondence between the languages of $(ab^\star) \cap a$ and $(a_1 b_2^\star) \cap a_3$, as it is the case for expressions without intersection. However, the various conversions from expressions to automata based on the notions of derivative or partial derivative can still be extended to regular expressions with intersection [7, 9, 2, 3]. In this paper, we present a position automaton construction for regular expressions with intersection by generalizing the notion of position. Instead of positions, sets of positions are considered, such that marking a regular expression is made compatible with the intersection operation. We also show that the partial derivative automaton is a quotient of this new position automaton. A preliminary and shorter version of this paper was previously published [5].

The rest of the paper is organised as follows. Section 2 recalls some basic notions on semi-extended regular expressions and finite automata, and presents a partial derivative automaton for those expressions, $\mathcal{A}_{\mathsf{pd}}$. In Section 3, the notions of indexed regular expressions, well-indexed regular expressions and indexed languages are introduced. Based on these concepts, the position automaton for semi-extended expressions, $\mathcal{A}_{\mathsf{pos}}$, is presented in Section 4. To calculate $\mathcal{A}_{\mathsf{pos}}$ we give recursive definitions of supersets of *first*, *last* and *follow* sets. It is proved that the trimmed version of the resulting automaton is $\mathcal{A}_{\mathsf{pos}}$. For simple regular expressions, the position automaton is isomorphic to a *continuation* automaton used to prove that it has $\mathcal{A}_{\mathsf{pd}}$ as a quotient. Following a similar path, in Section 5 we define a c-continuation automaton for semi-extended expressions and in Section 6 we show that $\mathcal{A}_{\mathsf{pd}}$ is a quotient of $\mathcal{A}_{\mathsf{pos}}$. Section 7 concludes the paper with some final remarks.

2. Preliminaries

In this section we recall the basic definitions to be used throughout this paper and the notation. For further details we refer to [15, 19].

2.1. Regular Expressions, Languages and Finite Automata

Let Σ be an *alphabet* (set of letters). A *word* over Σ is a finite sequence of letters, where ε is the empty word. The size of a word x, $|x|$, is the number of alphabet symbols in x. $\Sigma^\star$ denotes the set of all words over Σ, and a *language* over Σ is any subset of $\Sigma^\star$. If $x = uv$ then v is a *suffix* of x and let $\mathsf{suff}(x)$ denote the language of all suffixes of x. The *concatenation* of two languages L_1 and L_2 is defined by $L_1 \cdot L_2 = \{\, xy \mid x \in L_1, y \in L_2 \,\}$, and $L^\star$ denotes the set $\{\, x_1 x_2 \cdots x_n \mid n \geq 0,\ x_i \in L \,\}$. The *left quotient* of a language $L \subseteq \Sigma^\star$ w.r.t. a word $x \in \Sigma^\star$ is the language $x^{-1}L = \{\, y \mid xy \in L \,\}$.

The set $\mathsf{RE}_\cap$ of *regular expressions with intersection* or *semi-extended expressions* over Σ is defined by the following grammar

$$\alpha, \beta := \emptyset \mid \varepsilon \mid a \in \Sigma \mid (\alpha + \beta) \mid (\alpha \cap \beta) \mid (\alpha \cdot \beta) \mid (\alpha^\star), \tag{1}$$

where the concatenation operator $\cdot$ is often omitted. We consider $\mathsf{RE}_\cap$ expressions modulo the standard equations for $\emptyset$ and ε, i.e. $\alpha + \emptyset = \emptyset + \alpha = \alpha \cdot \varepsilon = \varepsilon \cdot \alpha = \alpha$, $\alpha \cdot \emptyset = \emptyset \cdot \alpha = \alpha \cap \emptyset = \emptyset \cap \alpha = \emptyset$, and $\emptyset^\star = \varepsilon$. Throughout this paper we often refer to regular expressions with intersection just as regular expressions. The set of alphabet symbols with occurrences in α is denoted by Σ_α. Expressions containing no occurrence of the operator $\cap$ are called *simple regular expressions*. A *linear regular expression* is a regular expression in which every alphabet symbol occurs at most once. We let $|\alpha|$, $|\alpha|_\Sigma$ and $|\alpha|_\cap$ denote for $\alpha \in \mathsf{RE}_\cap$ the number of symbols (size), the number of occurrences of alphabet symbols and the number of occurrences of the binary operator $\cap$, respectively.

Definition 1. The language $\mathcal{L}(\alpha)$ associated to an expression $\alpha \in \mathsf{RE}_\cap$ is inductively defined as follows.

$$\begin{aligned} \mathcal{L}(\emptyset) &= \emptyset, & \mathcal{L}(\alpha \cdot \beta) &= \mathcal{L}(\alpha) \cdot \mathcal{L}(\beta), \\ \mathcal{L}(\varepsilon) &= \{\varepsilon\}, & \mathcal{L}(\alpha + \beta) &= \mathcal{L}(\alpha) \cup \mathcal{L}(\beta), \\ \mathcal{L}(a) &= \{a\}, & \mathcal{L}(\alpha \cap \beta) &= \mathcal{L}(\alpha) \cap \mathcal{L}(\beta), \\ \mathcal{L}(\alpha^\star) &= \mathcal{L}(\alpha)^\star. \end{aligned}$$

The language of $S \subseteq \mathsf{RE}_\cap$ is $\mathcal{L}(S) = \cup_{\alpha \in S} \mathcal{L}(\alpha)$. Given an expression $\alpha \in \mathsf{RE}_\cap$, we define $\varepsilon(\alpha) = \varepsilon$ if $\varepsilon \in \mathcal{L}(\alpha)$, and $\varepsilon(\alpha) = \emptyset$ otherwise. A recursive definition of $\varepsilon : \mathsf{RE}_\cap \longrightarrow \{\emptyset, \varepsilon\}$ is given by the following: $\varepsilon(a) = \varepsilon(\emptyset) = \emptyset$, $\varepsilon(\varepsilon) = \varepsilon(\alpha^\star) = \varepsilon$, $\varepsilon(\alpha + \beta) = \varepsilon(\alpha) + \varepsilon(\beta)$, and $\varepsilon(\alpha\beta) = \varepsilon(\alpha \cap \beta) = \varepsilon(\alpha) \cdot \varepsilon(\beta)$.

A *nondeterministic finite automaton* (NFA) is a tuple $\mathcal{A} = \langle S, \Sigma, S_0, \delta, F \rangle$, where S is a finite set of states, Σ is a finite alphabet, $S_0 \subseteq S$ a set of initial states, $\delta : S \times \Sigma \longrightarrow \mathcal{P}(S)$ the transition function, and $F \subseteq S$ a set of final states. The extension of δ to sets of states and words is defined by $\delta(X, \varepsilon) = X$ and $\delta(X, ax) = \delta(\cup_{s \in X} \delta(s, a), x)$. A word $x \in \Sigma^\star$ is accepted by $\mathcal{A}$ if and only if $\delta(S_0, x) \cap F \neq \emptyset$. The *language of* $\mathcal{A}$, $\mathcal{L}(\mathcal{A})$, is the set of words accepted by $\mathcal{A}$. The *right language of a state* s, $\mathcal{L}_s$, is the language accepted by $\mathcal{A}$ if $S_0 = \{s\}$. Two automata are *equivalent* if they accept the same language. If two automata $\mathcal{A}$ and $\mathcal{B}$ are isomorphic, we write $\mathcal{A} \simeq \mathcal{B}$.

An NFA is *initially connected* or *accessible* if each state is reachable from an initial state and it is *trimmed* if, moreover, the right language of each state is non-empty. Given $\mathcal{A}$, we denote by $\mathcal{A}^{\mathsf{ac}}$ and $\mathcal{A}^{\mathsf{t}}$ the result of removing unreachable states from $\mathcal{A}$ and trimming $\mathcal{A}$, respectively. It is clear that $\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}^{\mathsf{ac}}) = \mathcal{L}(\mathcal{A}^{\mathsf{t}})$.

An equivalence relation $\equiv$ over S is right invariant w.r.t. $\mathcal{A}$ iff

(I) $\forall s, t \in S,\ s \equiv t \wedge s \in F \implies t \in F$;

(II) $\forall s, t \in S, \forall a \in \Sigma,\ s \equiv t \implies \forall s_1 \in \delta(s, a)\ \exists t_1 \in \delta(t, a), s_1 \equiv t_1$.

If $\equiv$ is right invariant, then we can define the quotient automaton $\mathcal{A}/_\equiv$ in the usual way, and $\mathcal{L}(\mathcal{A}/_\equiv) = \mathcal{L}(\mathcal{A})$.

A standard conversion from a simple regular expression α to an equivalent NFA is the position/Glushkov automaton. Given a simple regular expression α, one can mark each occurrence of a letter with its position in α, considering reading it from left to right. The resulting regular expression is a *marked* regular expression $\overline{\alpha}$ with all symbols distinct and over an alphabet denoted by $\Sigma_{\overline{\alpha}}$. Then, a position $i \in [1, |\alpha|_\Sigma]$ corresponds to the symbol a_i in $\overline{\alpha}$, and consequently to exactly one occurrence of a in α. Let $\mathsf{pos}(\alpha) = \{1, 2, \ldots, |\alpha|_\Sigma\}$ and let $\mathsf{pos}_0(\alpha) = \mathsf{pos}(\alpha) \cup \{0\}$. For α and $i \in \mathsf{pos}(\alpha)$, let the sets *first*, *last* and *follow* be $\mathsf{First}(\alpha) = \{\ i \mid \exists w \in \Sigma^\star_{\overline{\alpha}}, \sigma_i w \in \mathcal{L}(\overline{\alpha})\ \}$, $\mathsf{Last}(\alpha) = \{\ i \mid \exists w \in \Sigma^\star_{\overline{\alpha}}, w\sigma_i \in \mathcal{L}(\overline{\alpha})\ \}$ and $\mathsf{Follow}(\alpha, i) = \{\ j \mid \exists u, v \in \Sigma^\star_{\overline{\alpha}}, u\sigma_i\sigma_j v \in \mathcal{L}(\overline{\alpha})\ \}$, respectively. The *position/Glushkov automaton* for a simple regular expression α is $\mathcal{A}_{\mathsf{pos}}(\alpha) = \langle \mathsf{pos}_0(\alpha), \Sigma, 0, \delta_{\mathsf{pos}}, F\rangle$, with $\delta_{\mathsf{pos}} = \{\ (0, \overline{a}_j, j) \mid j \in \mathsf{First}(\alpha)\ \} \cup \{\ (i, \overline{a}_j, j) \mid\ j \in \mathsf{Follow}(\alpha, i)\ \}$ and $F = \mathsf{Last}(\alpha) \cup \{0\}$ if $\varepsilon(\alpha) = \varepsilon$, and $F = \mathsf{Last}(\alpha)$, otherwise.

2.2. Partial Derivatives and the Partial Derivative Automaton

The notions of partial derivatives and partial derivative automata were introduced by Antimirov [1] for simple regular expressions. Bastos et al. [2, 3] presented an extension of the Antimirov construction from $\mathsf{RE}_\cap$ expressions.

Definition 2. For $\alpha \in \mathsf{RE}_\cap$ and $a \in \Sigma$, the set $\partial_a(\alpha)$ of *partial derivatives* of α w.r.t. a is defined by:

$$\partial_a(\emptyset) = \partial_a(\varepsilon) = \emptyset, \qquad \partial_a(\alpha + \beta) = \partial_a(\alpha) \cup \partial_a(\beta),$$

$$\partial_a(b) = \begin{cases} \{\varepsilon\}, & \text{if } a = b \\ \emptyset & \text{otherwise,} \end{cases} \qquad \partial_a(\alpha\beta) = \begin{cases} (\partial_a(\alpha) \odot \beta) \cup \partial_a(\beta), & \text{if } \varepsilon(\alpha) = \varepsilon \\ \partial_a(\alpha) \odot \beta, & \text{otherwise,} \end{cases}$$

$$\partial_a(\alpha^\star) = \partial_a(\alpha) \odot \alpha^\star, \qquad \partial_a(\alpha \cap \beta) = \partial_a(\alpha) \Cap \partial_a(\beta),$$

where for $S, T \subseteq \mathsf{RE}_\cap$ and $\beta \in \mathsf{RE}_\cap$, $S \odot \beta = \{\ \alpha\beta \mid \alpha \in S\ \}$, $\beta \odot S = \{\ \beta\alpha \mid \alpha \in S\ \}$, and $S \Cap T = \{\ \alpha \cap \beta \mid \alpha \in S, \beta \in T\ \}$.

This definition is extended to any word w by $\partial_\varepsilon(\alpha) = \{\alpha\}$, $\partial_{wa}(\alpha) = \bigcup_{\alpha_i \in \partial_w(\alpha)} \partial_a(\alpha_i)$, and $\partial_w(R) = \bigcup_{\alpha_i \in R} \partial_w(\alpha_i)$, where $R \subseteq \mathsf{RE}_\cap$. The set of partial derivatives of an expression α is $\partial(\alpha) = \bigcup_{w \in \Sigma^\star} \partial_w(\alpha)$. As for simple regular expressions, the partial derivative automaton of an expression $\alpha \in \mathsf{RE}_\cap$ is defined by $\mathcal{A}_{\mathsf{pd}}(\alpha) = \langle \partial(\alpha), \Sigma, \{\alpha\}, \delta_{\mathsf{pd}}, F_{\mathsf{pd}}\rangle$, where $F_{\mathsf{pd}} = \{\ \gamma \in \partial(\alpha) \mid \varepsilon(\gamma) = \varepsilon\ \}$ and $\delta_{\mathsf{pd}}(\gamma, a) = \partial_a(\gamma)$.

It follows that $\mathcal{L}(\mathcal{A}_{\mathsf{pd}}(\alpha))$ is exactly $\mathcal{L}(\alpha)$ and by construction $\mathcal{A}_{\mathsf{pd}}(\alpha)$ is accessible. An illustrative example can be found in Figure 2. Bastos et al. showed also that $|\partial(\alpha)| \leq 2^{|\alpha|_\Sigma - |\alpha|_\cap - 1} + 1$ and on average an asymptotical upper bound for the number of states is $(1.056 + o(1))^n$, where n is the size of the expression.

3. Indexed Expressions

Given an alphabet Σ and a nonempty set of indexes $J \subseteq \mathbb{N}$, let $\Sigma_J = \{\ a_j \mid a \in \Sigma, j \in J\ \}$. An *indexed regular expression* is a regular expression over the alphabet Σ_J such that for all $a_i, b_j \in \Sigma_J$ occurring in the expression, $a \neq b$ implies $i \neq j$. We let $\rho, \rho_1, \rho_2, \ldots$ denote indexed regular expressions. If ρ is an indexed expression, then $\overline{\rho}$ is the regular expression over the alphabet Σ obtained by removing the indexes. The set of all indexes occurring in ρ is denoted by $\mathsf{ind}(\rho) = \{\ i \mid a_i \in \Sigma_\rho\ \}$. Given an indexed expression ρ and $i \in \mathsf{ind}(\rho)$, $\ell_\rho(i)$ is the (unindexed) letter indexed by i in ρ. From now on, we will simply write $\ell(i)$ for $\ell_\rho(i)$ since it will always be clear that we are referring to a specific expression ρ. Given an indexed expression ρ, let

$$\mathcal{I}_\rho = \{\ \mathrm{I} \subseteq \mathsf{ind}(\rho) \mid \mathrm{I} \neq \emptyset \text{ and } \forall i_1, i_2 \in \mathrm{I}, \ell(i_1) = \ell(i_2)\ \}.$$

For $\mathrm{I} \in \mathcal{I}_\rho$, the definition of ℓ is extended to $\ell(\mathrm{I}) = \ell(i)$, $i \in \mathrm{I}$. Finally, one says that ρ is *well-indexed* if for all subterms of ρ of the form $\rho_1 \cap \rho_2$ one has $\mathsf{ind}(\rho_1) \cap \mathsf{ind}(\rho_2) = \emptyset$.

Example 3. For $\rho = a_1(a_4 b_5^\star \cap a_4)$ one has $\overline{\rho} = a(ab^\star \cap a)$, $\mathsf{ind}(\rho) = \{1, 4, 5\}$, $\ell(4) = \ell(\{1, 4\}) = a$ and $\mathcal{I}_\rho = \{\{1\}, \{4\}, \{5\}, \{1, 4\}\}$. However, this expression is not well-indexed, since a_4 occurs on both sides of an intersection.

Definition 4. Consider an indexed expression ρ. For $L \subseteq \mathcal{I}_\rho^\star$ and $x = \mathrm{I}_1 \cdots \mathrm{I}_n \in L$, we define $\ell(x) = \ell(\mathrm{I}_1) \cdots \ell(\mathrm{I}_n)$ and $\ell(L) = \{\ \ell(x) \mid x \in L\ \}$. The *indexed intersection* of two words $x = \mathrm{I}_1 \cdots \mathrm{I}_m, y = \mathrm{J}_1 \cdots \mathrm{J}_n \in \mathcal{I}_\rho^\star$ is defined by $x \cap_\mathcal{I} y = (\mathrm{I}_1 \cup \mathrm{J}_1) \cdots (\mathrm{I}_n \cup \mathrm{J}_n)$ if $\ell(x) = \ell(y)$[1], and undefined otherwise. Then, the *indexed intersection* of two languages $L_1, L_2 \in \mathcal{I}_\rho^\star$ is defined as follows:

$$L_1 \cap_\mathcal{I} L_2 = \{\ x \cap_\mathcal{I} y \mid x \in L_1, y \in L_2\ \}.$$

The *index-language* $\mathcal{L}_\mathcal{I}(\rho) \subseteq \mathcal{I}_\rho^\star$ associated with ρ is defined as follows.

$$\begin{aligned}
\mathcal{L}_\mathcal{I}(\emptyset) &= \emptyset, & \mathcal{L}_\mathcal{I}(\varepsilon) &= \{\varepsilon\},\\
\mathcal{L}_\mathcal{I}(\rho^\star) &= \mathcal{L}_\mathcal{I}(\rho)^\star, & \mathcal{L}_\mathcal{I}(a_i) &= \{\{i\}\},\\
\mathcal{L}_\mathcal{I}(\rho_1 + \rho_2) &= \mathcal{L}_\mathcal{I}(\rho_1) \cup \mathcal{L}_\mathcal{I}(\rho_2), & \mathcal{L}_\mathcal{I}(\rho_1 \cdot \rho_2) &= \mathcal{L}_\mathcal{I}(\rho_1) \cdot \mathcal{L}_\mathcal{I}(\rho_2),\\
\mathcal{L}_\mathcal{I}(\rho_1 \cap \rho_2) &= \mathcal{L}_\mathcal{I}(\rho_1) \cap_\mathcal{I} \mathcal{L}_\mathcal{I}(\rho_2).
\end{aligned}$$

Example 5. For $\rho = (a_1 a_2 + b_3 + a_4)^\star \cap (a_5 + b_6)^\star$, we have $\mathcal{L}_\mathcal{I}(\rho) = \{\{4, 5\}, \{3, 6\}, \{1, 5\}\{2, 5\}, \{4, 5\}\{4, 5\}, \{4, 5\}\{3, 6\}, \ldots\}$, and $\ell(\mathcal{L}_\mathcal{I}(\rho)) = \{a, b, aa, ab, \ldots\}$ (since $\ell(\{1, 5\}\{2, 5\}) = \ell(\{4, 5\}\{4, 5\}) = aa$).

Proposition 6. *Given an indexed expression ρ, one has $\ell(\mathcal{L}_\mathcal{I}(\rho)) = \mathcal{L}(\overline{\rho})$.*

Proof. It is easy to show by induction on the structure of ρ that $x \in \mathcal{L}_\mathcal{I}(\rho)$ implies $\ell(x) \in \mathcal{L}(\overline{\rho})$, and that for every $y \in \mathcal{L}(\overline{\rho})$ there is some $x \in \mathcal{L}_\mathcal{I}(\rho)$ such that $\ell(x) = y$. □

4. A Position Automaton for $\mathsf{RE}_\cap$ Expressions

Given $\alpha \in \mathsf{RE}_\cap$, the indexed expression $\overline{\alpha}$ is always linear (thus well-indexed), and also $\mathsf{pos}(\alpha) = \mathsf{ind}(\overline{\alpha})$. For an indexed linear expression ρ, we define the following

[1]Note that $\ell(x) = \ell(y)$ implies that $m = n$ and that $\ell(x \cap_\mathcal{I} y) = \ell(x) = \ell(y)$.

subsets of $\mathcal{I}_\rho$:

$$\begin{aligned}\mathsf{First}'(\rho) &= \{\ \mathrm{I} \mid \exists x,\ \mathrm{I}x \in \mathcal{L}_{\mathcal{I}}(\rho)\ \},\\ \mathsf{Last}'(\rho) &= \{\ \mathrm{I} \mid \exists x,\ x\mathrm{I} \in \mathcal{L}_{\mathcal{I}}(\rho)\ \},\\ \mathsf{Follow}'(\rho) &= \{\ (\mathrm{I},\mathrm{J}) \mid \exists x,y,\ x\mathrm{I}\mathrm{J}y \in \mathcal{L}_{\mathcal{I}}(\rho)\ \}.\end{aligned}$$

Then, given $\alpha \in \mathsf{RE}_\cap$, we define $\mathsf{First}(\alpha) = \mathsf{First}'(\overline{\alpha})$, $\mathsf{Last}(\alpha) = \mathsf{Last}'(\overline{\alpha})$, and $\mathsf{Follow}(\alpha) = \mathsf{Follow}'(\overline{\alpha})$.

Definition 7. The *position automaton* of an expression $\alpha \in \mathsf{RE}_\cap$ is

$$\mathcal{A}_{\mathsf{pos}}(\alpha) = \langle S_{\mathsf{pos}}, \Sigma, \{\{0\}\}, \delta_{\mathsf{pos}}, F_{\mathsf{pos}} \rangle,$$

where

$$\begin{aligned}S_{\mathsf{pos}} &= \{\{0\}\} \cup \{\ \mathrm{I} \in \mathcal{I}_{\overline{\alpha}} \mid x\mathrm{I}y \in \mathcal{L}_{\mathcal{I}}(\overline{\alpha}) \text{ for some } x,y \in \mathcal{I}^\star_{\overline{\alpha}}\ \},\\ \delta_{\mathsf{pos}} &= \{\ (\mathrm{I},\ell(\mathrm{J}),\mathrm{J}) \mid (\mathrm{I},\mathrm{J}) \in \mathsf{Follow}(\alpha)\ \} \cup \{\ (\{0\},\ell(\mathrm{I}),\mathrm{I}) \mid \mathrm{I} \in \mathsf{First}(\alpha)\ \},\\ F_{\mathsf{pos}} &= \begin{cases}\mathsf{Last}(\alpha) \cup \{\{0\}\}, & \text{if } \varepsilon(\alpha) = \varepsilon;\\ \mathsf{Last}(\alpha), & \text{otherwise.}\end{cases}\end{aligned}$$

The following proposition is a consequence of Proposition 6 and the corresponding result for simple regular expressions [14].

Proposition 8. *Given an expression $\alpha \in \mathsf{RE}_\cap$, one has $\mathcal{L}(\mathcal{A}_{\mathsf{pos}}(\alpha)) = \mathcal{L}(\alpha)$.*

Note that for regular expressions without intersection (simple regular expressions) the automaton is, by the definition of $\mathcal{L}_{\mathcal{I}}$, isomorphic to the classic position automaton, with the difference that now states are labelled with singletons $\{i\}$ instead of $i \in \mathsf{pos}(\alpha) \cup \{0\}$.

4.1. Recursive Definitions

We now give definitions for recursively computing sets corresponding to First, Last and Follow. These definitions lead to supersets of the corresponding sets but we will prove that extra elements can be discarded and if we trim the resulting NFA we obtain $\mathcal{A}_{\mathsf{pos}}$. Again, considering simple regular expressions these recursive definitions coincide with the ones for the sets First, Last and Follow.

Definition 9. Given a well-indexed expression ρ, let $\mathsf{Fst}(\rho) \subseteq \mathcal{I}_\rho$ be inductively defined as follows,

$$\begin{aligned}&\mathsf{Fst}(\emptyset) = \mathsf{Fst}(\varepsilon) = \emptyset, && \mathsf{Fst}(\rho_1 + \rho_2) = \mathsf{Fst}(\rho_1) \cup \mathsf{Fst}(\rho_2),\\ &\mathsf{Fst}(a_i) = \{\{i\}\}, && \mathsf{Fst}(\rho_1 \cdot \rho_2) = \begin{cases}\mathsf{Fst}(\rho_1) \cup \mathsf{Fst}(\rho_2), & \text{if } \varepsilon(\rho_1) = \varepsilon;\\ \mathsf{Fst}(\rho_1), & \text{otherwise,}\end{cases}\\ &\mathsf{Fst}(\rho^\star) = \mathsf{Fst}(\rho), && \mathsf{Fst}(\rho_1 \cap \rho_2) = \mathsf{Fst}(\rho_1) \otimes \mathsf{Fst}(\rho_2),\end{aligned}$$

where for $F_1, F_2 \subseteq \mathcal{I}_\rho$, $F_1 \otimes F_2 = \{\ \mathrm{I}_1 \cup \mathrm{I}_2 \mid \ell(\mathrm{I}_1) = \ell(\mathrm{I}_2) \text{ and } \mathrm{I}_1 \in F_1,\ \mathrm{I}_2 \in F_2\ \}$.

By construction, all elements $\mathrm{I} \in \mathsf{Fst}(\rho)$ are non-empty and such that $\ell(i_1) = \ell(i_2)$ for all $i_1, i_2 \in \mathrm{I}$, guaranting that $\otimes$ is well defined and $\mathsf{Fst}(\rho) \subseteq \mathcal{I}_\rho$.

Example 10. We have $\mathsf{Fst}(a_1^\star b_2^\star \cap a_3) = \mathsf{Fst}(a_1^\star b_2^\star) \otimes \mathsf{Fst}(a_3) = \{\{1\},\{2\}\} \otimes \{\{3\}\} = \{\{1,3\}\}$.

Definition 11. Given a well-indexed expression ρ, the set $\mathsf{Lst}(\rho) \subseteq \mathcal{I}_\rho$ is defined as $\mathsf{Fst}(\rho)$, with the difference that for concatenation we have:

$$\mathsf{Lst}(\rho_1 \cdot \rho_2) = \begin{cases} \mathsf{Lst}(\rho_1) \cup \mathsf{Lst}(\rho_2), & \text{if } \varepsilon(\rho_2) = \varepsilon; \\ \mathsf{Lst}(\rho_2), & \text{otherwise.} \end{cases}$$

The set $\mathsf{Fol}(\rho) \subseteq \mathcal{I}_\rho \times \mathcal{I}_\rho$ is inductively defined as follows,

$$\begin{aligned} \mathsf{Fol}(\emptyset) &= \mathsf{Fol}(\varepsilon) = \mathsf{Fol}(a_i) = \emptyset & \mathsf{Fol}(\rho_1 + \rho_2) &= \mathsf{Fol}(\rho_1) \cup \mathsf{Fol}(\rho_2) \\ \mathsf{Fol}(\rho^\star) &= \mathsf{Fol}(\rho) \cup \mathsf{Lst}(\rho) \times \mathsf{Fst}(\rho) & \mathsf{Fol}(\rho_1 \cap \rho_2) &= \mathsf{Fol}(\rho_1) \otimes \mathsf{Fol}(\rho_2) \end{aligned}$$
$$\mathsf{Fol}(\rho_1 \cdot \rho_2) = \mathsf{Fol}(\rho_1) \cup \mathsf{Fol}(\rho_2) \cup \mathsf{Lst}(\rho_1) \times \mathsf{Fst}(\rho_2).$$

where, for $S_1, S_2 \subseteq \mathcal{I}_\rho \times \mathcal{I}_\rho$,

$$S_1 \otimes S_2 = \{\ (\mathrm{I}_1 \cup \mathrm{I}_2, \mathrm{J}_1 \cup \mathrm{J}_2) \mid (\mathrm{I}_1, \mathrm{J}_1) \in S_1, (\mathrm{I}_2, \mathrm{J}_2) \in S_2 \text{ and } \ell(\mathrm{I}_1) = \ell(\mathrm{I}_2), \ell(\mathrm{J}_1) = \ell(\mathrm{J}_2)\ \}.$$

In the next definition we will use the standard projection functions on the first and second coordinates, π_1 and π_2, respectively.

Definition 12. Given $\alpha \in \mathsf{RE}_\cap$, let $\mathcal{A}_{\mathsf{posi}}(\alpha) = \langle S_{\mathsf{posi}}, \Sigma, \{\{0\}\}, \delta_{\mathsf{posi}}, F_{\mathsf{posi}} \rangle$ be the NFA where $S_{\mathsf{posi}} = \{\{0\}\} \cup \mathsf{Fst}(\overline{\alpha}) \cup \mathsf{Lst}(\overline{\alpha}) \cup \pi_1(\mathsf{Fol}(\overline{\alpha})) \cup \pi_2(\mathsf{Fol}(\overline{\alpha}))$, and δ_{posi} and F_{posi} are defined as δ_{pos} and F_{pos}, in Definition 7 , substituting the functions First, Last and Follow, by Fst, Lst and Fol, respectively.

We will now show that $\mathcal{L}(\mathcal{A}_{\mathsf{pos}}(\alpha)) = \mathcal{L}(\mathcal{A}_{\mathsf{posi}}(\alpha))$, and that $\mathcal{A}_{\mathsf{pos}}(\alpha)$ is obtained by trimming $\mathcal{A}_{\mathsf{posi}}(\alpha)$, as the result of the two following lemmas. An example is presented at the end of this section. The first lemma ensures that $\mathcal{L}(\mathcal{A}_{\mathsf{pos}}(\alpha)) \subseteq \mathcal{L}(\mathcal{A}_{\mathsf{posi}}(\alpha))$.

Lemma 13. *Given an indexed linear expression ρ, one has:*
- (I) $\mathsf{First}'(\rho) \subseteq \mathsf{Fst}(\rho)$*;*
- (II) $\mathsf{Last}'(\rho) \subseteq \mathsf{Lst}(\rho)$*;*
- (III) $\mathsf{Follow}'(\rho) \subseteq \mathsf{Fol}(\rho)$.

Proof. We proceed by induction on the structure of the expression.
- (I) We only present the case of expressions of the form $\rho_1 \cap \rho_2$. If $\mathrm{I} \in \mathsf{First}'(\rho_1 \cap \rho_2)$, then there is $x \in \mathcal{I}^\star_{\rho_1 \cap \rho_2}$ such that $\mathrm{I}x \in \mathcal{L}_\mathcal{I}(\rho_1 \cap \rho_2)$. Thus, there exist $\mathrm{I}_1 x_1 \in \mathcal{L}_\mathcal{I}(\rho_1)$, $\mathrm{I}_2 x_2 \in \mathcal{L}_\mathcal{I}(\rho_2)$ such that $\mathrm{I} = \mathrm{I}_1 \cup \mathrm{I}_2$ and $x = x_1 \cap_\mathcal{I} x_2$. One has $\mathrm{I}_1 \in \mathsf{First}'(\rho_1)$ and $\mathrm{I}_2 \in \mathsf{First}'(\rho_2)$, and by the induction hypothesis, $\mathrm{I}_1 \in \mathsf{Fst}(\rho_1)$ and $\mathrm{I}_2 \in \mathsf{Fst}(\rho_2)$. By the definition of Fst we conclude that $\mathrm{I} \in \mathsf{Fst}(\rho_1 \cap \rho_2)$.

(II) This case is analogous to the previous one.

(III) We also only present the case of intersection. If $(\mathrm{I},\mathrm{J}) \in \mathsf{Follow}'(\rho_1 \cap \rho_2)$, then there are $x, y \in \mathcal{I}^\star_{\rho_1\cap\rho_2}$ such that $x\mathrm{IJ}y \in \mathcal{L}_\mathcal{I}(\rho_1 \cap \rho_2)$. Thus, there exist $x_1\mathrm{I}_1\mathrm{J}_1y_1 \in \mathcal{L}_\mathcal{I}(\rho_1)$ and $x_2\mathrm{I}_2\mathrm{J}_2y_2 \in \mathcal{L}_\mathcal{I}(\rho_2)$ such that $\mathrm{I} = \mathrm{I}_1 \cup \mathrm{I}_2$, $\mathrm{J} = \mathrm{J}_1 \cup \mathrm{J}_2$, $x = x_1 \cap_\mathcal{I} x_2$ and $y = y_1 \cap_\mathcal{I} y_2$. One has $(\mathrm{I}_1,\mathrm{J}_1) \in \mathsf{Follow}'(\rho_1)$ and $(\mathrm{I}_2,\mathrm{J}_2) \in \mathsf{Follow}'(\rho_2)$, and by the induction hypothesis, $(\mathrm{I}_1,\mathrm{J}_1) \in \mathsf{Fol}(\rho_1)$ and $(\mathrm{I}_2,\mathrm{J}_2) \in \mathsf{Fol}(\rho_2)$. By the definition of Fol we conclude that $(\mathrm{I},\mathrm{J}) \in \mathsf{Fol}(\rho_1 \cap \rho_2)$. □

The following example shows that in general the reverse inclusion does not hold.

Example 14. For $\rho = (a_1 \cap b_2)c_3d_4$, we have $(\{3\},\{4\}) \in \mathsf{Fol}(\rho)$, but $(\{3\},\{4\}) \notin \mathsf{Follow}(\rho)$. Thus, $\mathsf{Fol}(\rho) \nsubseteq \mathsf{Follow}'(\rho)$.

The previous lemma shows that for any $\alpha \in \mathsf{RE}_\cap$, $\mathcal{A}_{\mathsf{pos}}(\alpha)$ is a subautomaton of $\mathcal{A}_{\mathsf{posi}}(\alpha)$, and thus $\mathcal{L}(\mathcal{A}_{\mathsf{pos}}(\alpha)) \subseteq \mathcal{L}(\mathcal{A}_{\mathsf{posi}}(\alpha))$. The following lemma will be needed to show that both recognize the same language and can be made isomorphic by trimming $\mathcal{A}_{\mathsf{posi}}$.

Lemma 15. *Given an indexed linear expression ρ and some $n \geq 1$, if $\mathrm{I}_n \in \mathsf{Lst}(\rho)$ and there exist $\mathrm{I}_1,\ldots,\mathrm{I}_n \in \mathcal{I}_\rho$ such that*

$$(\{0\},\ell(\mathrm{I}_1),\mathrm{I}_1),(\mathrm{I}_1,\ell(\mathrm{I}_2),\mathrm{I}_2),\ldots,(\mathrm{I}_{n-1},\ell(\mathrm{I}_n),\mathrm{I}_n) \in \delta_{\mathsf{posi}},$$

then $\mathrm{I}_1\cdots\mathrm{I}_n \in \mathcal{L}_\mathcal{I}(\rho)$.

Proof. We proceed by induction on the structure of the expression. For the base cases of $\emptyset$, ε and a_i it is obvious.

Let ρ be of the form $\rho_1 + \rho_2$. Then, $\mathrm{I}_1 \in \mathsf{Fst}(\rho_1+\rho_2)$ and $(\mathrm{I}_j,\mathrm{I}_{j+1}) \in \mathsf{Fol}(\rho_1+\rho_2)$, for $1 \leq j \leq n-1$. Since ρ is linear, and therefore $\mathsf{ind}(\rho_1) \cap \mathsf{ind}(\rho_2) = \emptyset$, this implies that $\mathrm{I}_1 \in \mathsf{Fst}(\rho_i)$ and $(\mathrm{I}_j,\mathrm{I}_{j+1}) \in \mathsf{Fol}(\rho_i)$, for $1 \leq j \leq n-1$, where i is either 1 or 2. Also, $\mathrm{I}_n \in \mathsf{Lst}(\rho_i)$. Then, it follows from the induction hypothesis that $\mathrm{I}_1\cdots\mathrm{I}_n \in \mathcal{L}_\mathcal{I}(\rho_i) \subseteq \mathcal{L}_\mathcal{I}(\rho)$.

Let ρ be of the form $\rho_1\rho_2$. Then, $\mathrm{I}_1 \in \mathsf{Fst}(\rho_1\rho_2)$ and $(\mathrm{I}_j,\mathrm{I}_{j+1}) \in \mathsf{Fol}(\rho_1\rho_2)$, for $1 \leq j \leq n-1$. Since ρ is linear, there exists $l \in \{1,\ldots,n+1\}$ such that $\mathrm{I}_j \in \mathcal{I}_{\rho_1}$ for $1 \leq j \leq l-1$, $\mathrm{I}_j \in \mathcal{I}_{\rho_2}$ for $l \leq j \leq n+1$, and $(\mathrm{I}_{l-1},\mathrm{I}_l) \in \mathsf{Lst}(\rho_1)\times\mathsf{Fst}(\rho_2)$. By the same arguments as in the previous case, we conclude that $\mathrm{I}_1\cdots\mathrm{I}_{l-1} \in \mathcal{L}_\mathcal{I}(\rho_1)$, $\mathrm{I}_l\cdots\mathrm{I}_n \in \mathcal{L}_\mathcal{I}(\rho_2)$, thus $\mathrm{I}_1\cdots\mathrm{I}_n \in \mathcal{L}_\mathcal{I}(\rho_1)\mathcal{L}_\mathcal{I}(\rho_2) \subseteq \mathcal{L}_\mathcal{I}(\rho)$. Note that for $l=1$ (resp. $l=n+1$) the pair $(\mathrm{I}_{l-1},\mathrm{I}_l)$ does not exist and the whole sequence $\mathrm{I}_1\cdots\mathrm{I}_n$ is in $\mathcal{L}_\mathcal{I}(\rho_2)$ (resp. $\mathcal{L}_\mathcal{I}(\rho_1)$).

Let ρ be of the form $\rho_1^\star$. Then $\mathrm{I}_1 \in \mathsf{Fst}(\rho_1)$ and $(\mathrm{I}_j,\mathrm{I}_{j+1}) \in \mathsf{Fol}(\rho_1)\cup\mathsf{Lst}(\rho_1)\times\mathsf{Fst}(\rho_1)$, for $1 \leq j \leq n-1$. There exist $1 \leq k_1 < \cdots < k_m = n$, $m \geq 1$, such that $(\mathrm{I}_{k_i},\mathrm{I}_{k_i+1}) \in \mathsf{Lst}(\rho_1)\times\mathsf{Fst}(\rho_1)$, for $1 \leq i < m$, and for $j \neq k_i$, $(\mathrm{I}_j,\mathrm{I}_{j+1}) \in \mathsf{Fol}(\rho_1)$. It follows from the induction hypothesis that $\mathrm{I}_1\cdots\mathrm{I}_{k_1},\ldots,\mathrm{I}_{k_{m-1}}\cdots\mathrm{I}_{k_m} \in \mathcal{L}_\mathcal{I}(\rho_1)$. Thus, $\mathrm{I}_1\cdots\mathrm{I}_n \in \mathcal{L}_\mathcal{I}(\rho_1^\star) = \mathcal{L}_\mathcal{I}(\rho)$. If $k_1 = 1$ (resp. $k_m = n$) the pair $(\mathrm{I}_{k_i-1},\mathrm{I}_{k_i})$ does not exist but $\mathrm{I}_1 \in \mathsf{Fst}(\rho_1)$ (resp. $\mathrm{I}_n \in \mathsf{Lst}(\rho_1)$).

Let ρ be of the form $\rho_1 \cap \rho_2$. Then, $\mathrm{I}_1 \in \mathsf{Fst}(\rho_1\cap\rho_2)$ and $(\mathrm{I}_j,\mathrm{I}_{j+1}) \in \mathsf{Fol}(\rho_1\cap\rho_2)$, for $1 \leq j \leq n-1$. Since ρ is linear, we can write each I_j uniquely as $\mathrm{I}_j = \mathrm{I}_j^1 \cup \mathrm{I}_j^2$

with $\mathrm{I}_j^1 \subseteq \mathsf{ind}(\rho_1)$ and $\mathrm{I}_j^2 \subseteq \mathsf{ind}(\rho_2)$, for $1 \leq j \leq n$. Furthermore, $\mathrm{I}_1^k \in \mathsf{Fst}(\rho_k)$, $(\mathrm{I}_j^k, \mathrm{I}_{j+1}^k) \in \mathsf{Fol}(\rho_k)$, and $\mathrm{I}_n^k \in \mathsf{Lst}(\rho_k)$, for $1 \leq j \leq n-1$ and $k = 1, 2$. The result follows from the induction hypothesis. □

From the above, one has

Theorem 16. *For any* $\alpha \in \mathsf{RE}_\cap$, $\mathcal{L}(\mathcal{A}_{\mathsf{pos}}(\alpha)) = \mathcal{L}(\mathcal{A}_{\mathsf{posi}}(\alpha))$.

From these results, it follows that if we trim the automaton $\mathcal{A}_{\mathsf{posi}}$ we obtain exactly $\mathcal{A}_{\mathsf{pos}}$.

Corollary 17. $\mathcal{A}_{\mathsf{pos}}(\alpha) = \mathcal{A}_{\mathsf{posi}}(\alpha)^{\mathsf{t}}$.

Example 18. Consider $\alpha = (ba^\star b + a) \cap (aa + b)^\star$. Then $\overline{\alpha} = (b_1 a_2^\star b_3 + a_4) \cap (a_5 a_6 + b_7)^\star$, $\mathsf{Fst}(\overline{\alpha}) = \{\{1,7\},\{4,5\}\}$, $\mathsf{Lst}(\overline{\alpha}) = \{\{3,7\},\{4,6\}\}$, and $\mathsf{Fol}(\overline{\alpha}) = \{(\{2,5\},\{2,6\}),(\{2,6\},\{2,5\}),(\{2,6\},\{3,7\}),(\{1,7\},\{2,5\}),(\{1,7\},\{3,7\})\}$.

The automaton $\mathcal{A}_{\mathsf{posi}}(\alpha)$ is represented in Figure 1. The trimmed automaton, $\mathcal{A}_{\mathsf{posi}}(\alpha)^{\mathsf{t}}$, is obtained removing the states labeled by $\{4,5\}$ and $\{4,6\}$, and the corresponding transitions.

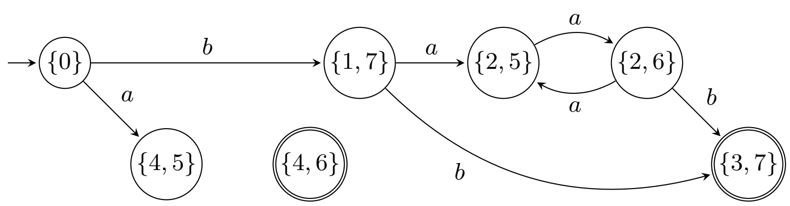

Figure 1: $\mathcal{A}_{\mathsf{posi}}((ba^\star b + a) \cap (aa + b)^\star)$

5. A c-Continuation Automaton for $\mathsf{RE}_\cap$ Expressions

In the case of simple regular expressions, Champarnaud and Ziadi [11] defined a nondeterministic automaton isomorphic to the position automaton, called the c-continuation automaton, in order to show that the partial derivative automaton can be seen as a quotient of the position automaton. With the same purpose, in this section, we present a c-continuation automaton for expressions with intersection. Moreover, instead of considering derivatives of regular expressions [7], we use partial derivatives to restate some known results for simple regular expressions.

5.1. Partial Index-Derivatives for $\mathsf{RE}_\cap$ *Expressions*

The notion of continuation was defined by Berry and Sethy [4], and developed by Champarnaud and Ziadi [11], by Ilie and Yu [16], and by Chen and Yu [12]. Given $a \in \Sigma$ and a linear simple expression α, the set of partial derivatives $\partial_{xa}(\alpha)$, for any word $x \in \Sigma^\star$, is either $\emptyset$ or has a unique element γ called the *continuation* of a in

α. Note that using partial derivatives, *continuations* and non-null c-*continuations* coincide. Furthermore, the continuation can be obtained by some refinement of the inductive definition of partial derivatives, exploring the linearity of α. In order to establish similar results for linear well-indexed expressions, we introduce the notion of *partial index-derivative* of a well-indexed expression ρ w.r.t. an index $\mathrm{I} \in \mathcal{I}_\rho$.

Given a well-indexed expression ρ, a subexpression τ of ρ, and a set of indexes $\mathrm{I} \in \mathcal{I}_\rho$, let $\mathrm{I}\big|_\tau$ denote the set of indexes in I that occur in τ. This definition is naturally extended to words $x = \mathrm{I}_1 \cdots \mathrm{I}_n \in \mathcal{I}_\rho^\star$ by $x\big|_\tau = \mathrm{I}_1\big|_\tau \cdots \mathrm{I}_n\big|_\tau$, for $n \geq 0$. In the next definitions, we use the operators $\odot$ and $\Cap$ defined in Definition 2.

Definition 19. The set of partial index-derivatives of a well-indexed expression ρ by $\mathrm{I} \in \mathcal{I}_\rho \cup \{\emptyset\}$, $\partial_\mathrm{I}(\rho)$, is defined by

$$\begin{aligned}
\partial_\mathrm{I}(\emptyset) &= \partial_\mathrm{I}(\varepsilon) = \emptyset,\\
\partial_\mathrm{I}(\rho^\star) &= \partial_\mathrm{I}(\rho) \odot \rho^\star,\\
\partial_\mathrm{I}(\rho_1 + \rho_2) &= \partial_\mathrm{I}(\rho_1) \cup \partial_\mathrm{I}(\rho_2),\\
\partial_\mathrm{I}(a_i) &= \begin{cases} \{\varepsilon\}, & \text{if } \mathrm{I} = \{i\};\\ \emptyset, & \text{otherwise,}\end{cases}\\
\partial_\mathrm{I}(\rho_1 \cdot \rho_2) &= \begin{cases} (\partial_\mathrm{I}(\rho_1) \odot \rho_2) \cup \partial_\mathrm{I}(\rho_2), & \text{if } \varepsilon(\rho_1) = \varepsilon;\\ \partial_\mathrm{I}(\rho_1) \odot \rho_2, & \text{otherwise,}\end{cases}\\
\partial_\mathrm{I}(\rho_1 \cap \rho_2) &= \begin{cases} \partial_{\mathrm{I}|_{\rho_1}}(\rho_1) \Cap \partial_{\mathrm{I}|_{\rho_2}}(\rho_2), & \text{if } \mathrm{I} = \mathrm{I}\big|_{\rho_1} \cup \mathrm{I}\big|_{\rho_2};\\ \emptyset, & \text{otherwise.}\end{cases}
\end{aligned}$$

The set of partial index-derivatives of ρ by a word $x \in \mathcal{I}_\rho^\star$ is then inductively defined by $\partial_\varepsilon(\rho) = \{\rho\}$ and $\partial_{x\mathrm{I}}(\rho) = \bigcup_{\rho' \in \partial_x(\rho)} \partial_\mathrm{I}(\rho')$. If S is a set of well-indexed expressions, $\partial_x(S) = \bigcup_{\rho \in S} \partial_x(\rho)$.

Example 20. We have $\partial_{\{1,3\}}(a_1^\star b_2^\star \cap a_3) = \partial_{\{1\}}(a_1^\star b_2^\star) \Cap \partial_{\{3\}}(a_3) = \{a_1^\star b_2^\star \cap \varepsilon\}$.

It is straightforward to see that $\partial_\emptyset(\rho) = \emptyset$ for all ρ. Although $\emptyset \notin \mathcal{I}_\rho$, the notion of partial index-derivative includes the derivative by an empty set of indexes, in order to guarantee that the derivative of an intersection is well-defined. Also note that the partial index-derivative of a well-indexed expression is still well-indexed.

Finally, the set of partial index-derivatives of ρ by all $\mathrm{I} \in \mathcal{I}_\rho$ can be calculated simultaneously using an extension of the linear form defined by Antimirov [1], i.e. considering pairs (I, ρ') where $\rho' \in \partial_\mathrm{I}(\rho)$. This form is suited for an efficient implementation, specially in the case of intersection as the sets of indexes to be considered are unions of the ones of the operands.

Definition 21. Given a well-indexed expression ρ, the linear form of ρ, $f(\rho)$, is defined inductively by

$$f(\emptyset) = f(\varepsilon) = \emptyset, \qquad f(a_i) = \{(\{i\}, \varepsilon)\},$$
$$f(\rho^\star) = f(\rho) \odot \rho^\star,$$
$$f(\rho_1 + \rho_2) = f(\rho_1) \cup f(\rho_2), \quad f(\rho_1\rho_2) = \begin{cases} (f(\rho_1) \odot \rho_2) \cup f(\rho_2) \text{ if } \varepsilon(\rho_1) = \varepsilon; \\ f(\rho_1) \odot \rho_2 \text{ otherwise,} \end{cases}$$

$$f(\rho_1 \cap \rho_2) = \{\ (\mathrm{I}_1 \cup \mathrm{I}_2, \rho_1' \cap \rho_2') \mid (\mathrm{I}_1, \rho_1') \in f(\rho_1), (\mathrm{I}_2, \rho_2') \in f(\rho_2),\ \ell(\mathrm{I}_1) = \ell(\mathrm{I}_2)\},$$

such that, as before, $\Gamma \odot \rho = \{\ (\mathrm{I}, \rho'\rho) \mid (\mathrm{I}, \rho') \in \Gamma\ \}$.

It easily follows that

$$\partial_{\mathrm{I}}(\rho) = \{\ \rho' \mid (\mathrm{I}, \rho') \in f(\rho)\ \}.$$

Example 22. Consider $\overline{\alpha} = (b_1 a_2^\star b_3 + a_4) \cap (a_5 a_6 + b_7)^\star$ as in Example 18. One has

$$\begin{aligned} f(b_1 a_2^\star b_3 + a_4) &= \{(\{1\}, a_2^\star b_3), (\{4\}, \varepsilon)\} \\ f((a_5a_6 + b_7)^\star) &= \{(\{5\}, a_6(a_5a_6 + b_7)^\star), (\{7\}, (a_5a_6 + b_7)^\star)\} \\ f(\overline{\alpha}) &= \{(\{1,7\}, a_2^\star b_3 \cap (a_5a_6 + b_7)^\star), (\{4,5\}, \varepsilon \cap a_6(a_5a_6 + b_7)^\star)\}. \end{aligned}$$

From which the partial derivatives $\partial_{\{1,7\}}(\overline{\alpha})$ and $\partial_{\{4,5\}}(\overline{\alpha})$ can be obtained.

The following lemma characterises some non-null partial derivatives and will be used in Proposition 26.

Lemma 23. *If $x = \mathrm{I}_1 \cdots \mathrm{I}_n$ and $\partial_x(\rho) \neq \emptyset$, then $x = x\big|_\rho$.*

Proof. The proof is trivial by induction on n. ☐

Proposition 24. *Consider a well-indexed expression ρ and $\mathrm{I} \in \mathcal{I}_\rho$. Then,*

$$\mathrm{I}^{-1}\mathcal{L}_{\mathcal{I}}(\rho) = \mathcal{L}_{\mathcal{I}}(\partial_{\mathrm{I}}(\rho)) \quad \textit{and} \quad \mathcal{L}_{\mathcal{I}}(\rho) = \mathcal{L}_{\mathcal{I}}\left(\bigcup_{\mathrm{I}\in\mathcal{I}_\rho} (\mathrm{I} \odot \partial_{\mathrm{I}}(\rho)) \cup \varepsilon(\rho)\right).$$

Proof. The proof of the first equality is by induction on the structure of ρ. We present the case of an expression of the form $\rho_1 \cap \rho_2$ and $\mathrm{I} = \mathrm{I}\big|_{\rho_1} \cup \mathrm{I}\big|_{\rho_2}$, $\mathrm{I}\big|_{\rho_1}, \mathrm{I}\big|_{\rho_2} \neq \emptyset$. First note that

$$\begin{aligned} (\mathrm{I}\big|_{\rho_1} \cup \mathrm{I}\big|_{\rho_2})^{-1}\mathcal{L}_{\mathcal{I}}(\rho_1 \cap \rho_2) &= (\mathrm{I}\big|_{\rho_1} \cup \mathrm{I}\big|_{\rho_2})^{-1}(\mathcal{L}_{\mathcal{I}}(\rho_1) \cap_{\mathcal{I}} \mathcal{L}_{\mathcal{I}}(\rho_2)) \\ &= \{\ x \cap_{\mathcal{I}} y \mid \mathrm{I}\big|_{\rho_1} x \in \mathcal{L}_{\mathcal{I}}(\rho_1), \mathrm{I}\big|_{\rho_2} y \in \mathcal{L}_{\mathcal{I}}(\rho_2)\ \} \\ &= \mathrm{I}\big|_{\rho_1}^{-1}\mathcal{L}_{\mathcal{I}}(\rho_1) \cap_{\mathcal{I}} \mathrm{I}\big|_{\rho_2}^{-1}\mathcal{L}_{\mathcal{I}}(\rho_2). \end{aligned}$$

Now,

$$\begin{aligned} \mathrm{I}^{-1}\mathcal{L}_{\mathcal{I}}(\rho_1 \cap \rho_2) &= \mathrm{I}\big|_{\rho_1}^{-1}\mathcal{L}_{\mathcal{I}}(\rho_1) \cap_{\mathcal{I}} \mathrm{I}\big|_{\rho_2}^{-1}\mathcal{L}_{\mathcal{I}}(\rho_2) \\ &= \mathcal{L}_{\mathcal{I}}(\partial_{\mathrm{I}|_{\rho_1}}(\rho_1)) \cap_{\mathcal{I}} \mathcal{L}_{\mathcal{I}}(\partial_{\mathrm{I}|_{\rho_2}}(\rho_2)) \\ &= \mathcal{L}_{\mathcal{I}}(\partial_{\mathrm{I}|_{\rho_1}}(\rho_1) \cap \partial_{\mathrm{I}|_{\rho_2}}(\rho_2)) \\ &= \mathcal{L}_{\mathcal{I}}(\partial_{\mathrm{I}}(\rho_1 \cap \rho_2)). \end{aligned}$$

The second equality follows trivially from the first one. □

Corollary 25. *For every well-indexed expression $\rho \in \mathsf{RE}_\cap$ and word $x \in \mathcal{I}_\rho^\star$, one has $x^{-1}\mathcal{L}_\mathcal{I}(\rho) = \mathcal{L}_\mathcal{I}(\partial_x(\rho))$ and $\mathcal{L}_\mathcal{I}(\rho) = \mathcal{L}_\mathcal{I}(\bigcup_{x\in\mathcal{I}_\rho^\star}(x \odot \partial_x(\rho)) \cup \varepsilon(\rho))$.*

Proof. The finiteness of the operands in the second equality follows from the fact that the derivatives for a regular language are finite [7]. □

The following is an adaptation, for partial index-derivatives and intersection, of a result due to Berry and Sethi [4].

Proposition 26. *Consider a linear indexed expression $\rho \in \mathsf{RE}_\cap$ and $x\mathrm{I} \in \mathcal{I}_\rho^\star$. The partial index-derivative $\partial_{x\mathrm{I}}(\rho)$ of ρ satisfies:*

$$\partial_{x\mathrm{I}}(\emptyset) = \partial_{x\mathrm{I}}(\varepsilon) = \emptyset,$$

$$\partial_{x\mathrm{I}}(a_i) = \begin{cases} \{\varepsilon\}, & \text{if } x\mathrm{I} = \{i\},\\ \emptyset, & \text{otherwise,}\end{cases}$$

$$\partial_{x\mathrm{I}}(\rho_1+\rho_2) = \begin{cases} \partial_{x\mathrm{I}}(\rho_1), & \text{if } x\mathrm{I} = (x\mathrm{I})\big|_{\rho_1},\\ \partial_{x\mathrm{I}}(\rho_2), & \text{if } x\mathrm{I} = (x\mathrm{I})\big|_{\rho_2},\\ \emptyset & \text{otherwise}\end{cases}$$

$$\partial_{x\mathrm{I}}(\rho_1\cdot\rho_2) = \begin{cases} \partial_{x\mathrm{I}}(\rho_1)\odot\rho_2, & \text{if } x\mathrm{I} = (x\mathrm{I})\big|_{\rho_1},\\ \partial_{z\mathrm{I}}(\rho_2), & \text{if } x = yz, \varepsilon(\partial_y(\rho_1)) = \varepsilon, z\mathrm{I} = (z\mathrm{I})\big|_{\rho_2},\\ \emptyset, & \text{otherwise,}\end{cases}$$

$$\partial_{x\mathrm{I}}(\rho^\star) \subseteq \bigcup_{v\in\mathsf{suff}(x)} \partial_{v\mathrm{I}}(\rho)\odot\rho^\star,$$

$$\partial_{x\mathrm{I}}(\rho_1\cap\rho_2) = \begin{cases} \partial_{(x\mathrm{I})|_{\rho_1}}(\rho_1) \Cap \partial_{(x\mathrm{I})|_{\rho_2}}(\rho_2), & \text{if } x\mathrm{I} = (x\mathrm{I})\big|_{\rho_1} \cap_\mathcal{I} (x\mathrm{I})\big|_{\rho_2},\\ \emptyset, & \text{otherwise.}\end{cases}$$

Proof. We proceed by induction on ρ. The cases $\emptyset$, ε, and a_i are trivial. Let ρ be $\rho_1+\rho_2$. We prove the result by induction on the length of $x \in \mathcal{I}_\rho^\star$. For $x = \varepsilon$, $\partial_\mathrm{I}(\rho_1+\rho_2) = \partial_\mathrm{I}(\rho_1)\cup\partial_\mathrm{I}(\rho_2)$. If $\partial_\mathrm{I}(\rho_1+\rho_2) \neq \emptyset$, then $\mathrm{I} = \mathrm{I}\big|_{\rho_1+\rho_2}$. Since ρ is linear, either $\mathrm{I} = \mathrm{I}\big|_{\rho_1}$ and $\mathrm{I}\big|_{\rho_2} = \emptyset$, or $\mathrm{I} = \mathrm{I}\big|_{\rho_2}$ and $\mathrm{I}\big|_{\rho_1} = \emptyset$. In the former case $\partial_\mathrm{I}(\rho_1+\rho_2) = \partial_\mathrm{I}(\rho_1)$ and in the latter $\partial_\mathrm{I}(\rho_1+\rho_2) = \partial_\mathrm{I}(\rho_2)$. For $x = y\mathrm{I}'$ and $\partial_{x\mathrm{I}}(\rho_1+\rho_2) \neq \emptyset$, $\partial_{x\mathrm{I}}(\rho_1+\rho_2) = \partial_\mathrm{I}(\partial_x(\rho_1+\rho_2)) = \partial_\mathrm{I}(\partial_x(\rho_i))$, and $x = x\big|_{\rho_i}$ for some $i \in \{1,2\}$ (otherwise, $\partial_x(\rho_1+\rho_2) = \emptyset$). Thus, $\partial_{x\mathrm{I}}(\rho_1+\rho_2) = \partial_{x\mathrm{I}}(\rho_i)$, for some $i\in\{1,2\}$.

Consider ρ as $\rho_1\cdot\rho_2$ and $x\mathrm{I} = \mathrm{I}_1\cdots\mathrm{I}_k\mathrm{I}_{k+1}\cdots\mathrm{I}_n$, where $n \geq 1$ and $0 \leq k \leq n$, such that $\mathrm{I}_1\cup\cdots\cup\mathrm{I}_k \subseteq \mathsf{ind}(\rho_1)$ and $\mathrm{I}_{k+1}\cup\cdots\cup\mathrm{I}_n \subseteq \mathsf{ind}(\rho_2)$. Then $\partial_{\mathrm{I}_1\cdots\mathrm{I}_n}(\rho_1\cdot\rho_2) = \partial_{\mathrm{I}_{k+1}\cdots\mathrm{I}_n}(\partial_{\mathrm{I}_1\cdots\mathrm{I}_k}(\rho_1)\odot\rho_2)$ because $\partial_{\mathrm{I}_i}(\rho_2) = \emptyset$ for $0 \leq i \leq k$. If $k = n$, then $\partial_{x\mathrm{I}}(\rho_1\cdot\rho_2) = \partial_{x\mathrm{I}}(\rho_1)\odot\rho_2$. Otherwise, either $\varepsilon(\partial_{\mathrm{I}_1\cdots\mathrm{I}_k}(\rho_1)) = \varepsilon$ and we have $\partial_{\mathrm{I}_1\cdots\mathrm{I}_n}(\rho_1\cdot\rho_2) = \partial_{\mathrm{I}_{k+1}\cdots\mathrm{I}_n}(\rho_2)$, and the second case follows, or $\partial_{\mathrm{I}_1\cdots\mathrm{I}_n}(\rho_1\cdot\rho_2) = \emptyset$.

Now consider $\rho = \rho_1^\star$. For $x = \varepsilon$, we have $\partial_{\mathrm{I}}(\rho_1^\star) = \partial_{\mathrm{I}}(\rho_1) \odot \rho_1^\star$. For $x = y\mathrm{I}'$,

$$\begin{aligned}
\partial_{x\mathrm{I}}(\rho_1^\star) &= \partial_{\mathrm{I}}(\partial_{y\mathrm{I}'}(\rho_1^\star)) \\
&\subseteq \partial_{\mathrm{I}}(\{\ \rho' \cdot \rho_1{}^\star \mid \rho' \in \partial_{v\mathrm{I}'}(\rho_1), v \in \mathsf{suff}(y)\ \}) \\
&\subseteq \bigcup_{v \in \mathsf{suff}(y)} \bigcup_{\rho' \in \partial_{v\mathrm{I}'}(\rho_1)} \partial_{\mathrm{I}}(\rho' \cdot \rho_1{}^\star) \ \cup\ \partial_{\mathrm{I}'\mathrm{I}}(\rho_1^\star) \\
&\subseteq \bigcup_{v \in \mathsf{suff}(y)} \partial_{\mathrm{I}}(\partial_{v\mathrm{I}'}(\rho_1)) \odot \rho_1^\star \cup \partial_{\mathrm{I}'\mathrm{I}}(\rho_1) \odot \rho_1^\star \\
&= \bigcup_{v \in \mathsf{suff}(x)} \partial_{v\mathrm{I}}(\rho_1) \odot \rho_1{}^\star.
\end{aligned}$$

We finally present the case of intersection. For $x = \varepsilon$ the result follows from the definition of index-derivative. Let $x = y\mathrm{I}'$ and suppose that $\partial_{x\mathrm{I}}(\rho_1 \cap \rho_2) \neq \emptyset$. Then, $x\mathrm{I} = (x\mathrm{I})\big|_{\rho_1 \cap \rho_2} = (x\big|_{\rho_1} \cap_{\mathcal{I}} x\big|_{\rho_2})(\mathrm{I}\big|_{\rho_1} \cup \mathrm{I}\big|_{\rho_2})$. Furthermore,

$$\begin{aligned}
\partial_{x\mathrm{I}}(\rho_1 \cap \rho_2) &= \partial_{\mathrm{I}}(\partial_x(\rho_1 \cap \rho_2)) = \partial_{\mathrm{I}}(\partial_{x|_{\rho_1}}(\rho_1) \Cap \partial_{x|_{\rho_2}}(\rho_2)) = \\
&= \bigcup_{\substack{\rho_i' \in \partial_{x|_{\rho_i}}(\rho_i) \\ \mathrm{I} = \mathrm{I}|_{\rho_1'} \cup \mathrm{I}|_{\rho_2'}}} \partial_{\mathrm{I}}(\rho_1' \cap \rho_2') = \bigcup_{\substack{\rho_i' \in \partial_{x|_{\rho_i}}(\rho_i) \\ \mathrm{I} = \mathrm{I}|_{\rho_1} \cup \mathrm{I}|_{\rho_2}}} \partial_{\mathrm{I}}(\rho_1' \cap \rho_2') = \\
&= \partial_{\mathrm{I}|_{\rho_1}}(\partial_{x|_{\rho_1}}(\rho_1)) \Cap \partial_{\mathrm{I}|_{\rho_2}}(\partial_{x|_{\rho_2}}(\rho_2)) = \partial_{(x\mathrm{I})|_{\rho_1}}(\rho_1) \Cap \partial_{(x\mathrm{I})|_{\rho_2}}(\rho_2).
\end{aligned}$$

For the fourth step note that $\mathsf{ind}(\rho_i') \subseteq \mathsf{ind}(\rho_i)$ $(i = 1, 2)$ and $\mathrm{I} = \mathrm{I}\big|_{\rho_1'} \cup \mathrm{I}\big|_{\rho_2'}$ imply that $\mathrm{I} = \mathrm{I}\big|_{\rho_1} \cup \mathrm{I}\big|_{\rho_2}$. Otherwise, if $\mathrm{I}\big|_{\rho_1'} \cup \mathrm{I}\big|_{\rho_2'} \subsetneq \mathrm{I}$ then, by Lemma 23, $\partial_{\mathrm{I}}(\rho_1' \cap \rho_2') = \emptyset$. □

The previous proposition implies that if $\partial_{x\mathrm{I}}(\rho) \neq \emptyset$, then it has only one element for every $x \in \mathcal{I}_\rho^\star$. This fact is proved in Proposition 28 and the unique element (if exists) is defined below.

Definition 27. Given a linear indexed expression ρ and a set of indexes I, the c-continuation $\mathsf{c}_{\mathrm{I}}(\rho)$ of ρ w.r.t. I is defined by the following rules.

$$\begin{aligned}
\mathsf{c}_{\mathrm{I}}(\emptyset) &= \mathsf{c}_{\mathrm{I}}(\varepsilon) = \emptyset, \\
\mathsf{c}_{\mathrm{I}}(\rho^\star) &= \mathsf{c}_{\mathrm{I}}(\rho)\rho^\star, \\
\mathsf{c}_{\mathrm{I}}(a_i) &= \begin{cases} \varepsilon, & \text{if } \mathrm{I} = \{i\}; \\ \emptyset, & \text{otherwise,} \end{cases} \\
\mathsf{c}_{\mathrm{I}}(\rho_1 + \rho_2) &= \begin{cases} \mathsf{c}_{\mathrm{I}}(\rho_1), & \text{if } \mathsf{c}_{\mathrm{I}}(\rho_1) \neq \emptyset; \\ \mathsf{c}_{\mathrm{I}}(\rho_2), & \text{otherwise,} \end{cases} \\
\mathsf{c}_{\mathrm{I}}(\rho_1 \cdot \rho_2) &= \begin{cases} \mathsf{c}_{\mathrm{I}}(\rho_1) \cdot \rho_2, & \text{if } \mathsf{c}_{\mathrm{I}}(\rho_1) \neq \emptyset; \\ \mathsf{c}_{\mathrm{I}}(\rho_2), & \text{otherwise,} \end{cases} \\
\mathsf{c}_{\mathrm{I}}(\rho_1 \cap \rho_2) &= \begin{cases} \mathsf{c}_{\mathrm{I}|_{\rho_1}}(\rho_1) \cap \mathsf{c}_{\mathrm{I}|_{\rho_2}}(\rho_2), & \text{if } \mathrm{I} = \mathrm{I}\big|_{\rho_1} \cup \mathrm{I}\big|_{\rho_1}; \\ \emptyset, & \text{otherwise.} \end{cases}
\end{aligned}$$

It is easy to verify that $\mathsf{c}_{\mathrm{I}}(\rho) \neq \emptyset$ implies $\mathrm{I} \subseteq \mathsf{ind}(\rho)$, i.e. $\mathrm{I}\big|_{\rho} = \mathrm{I}$.

Proposition 28. *Consider a linear indexed expression ρ and $\mathrm{I} \in \mathcal{I}_\rho$. Then, for every $x \in \mathcal{I}_\rho^\star$ such that $\partial_{x\mathrm{I}}(\rho) \neq \emptyset$, one has $\partial_{x\mathrm{I}}(\rho) = \{\mathsf{c}_{\mathrm{I}}(\rho)\}$ and $\mathsf{c}_{\mathrm{I}}(\rho) \neq \emptyset$.*

Proof. We proceed by induction on the structure of ρ. For $\emptyset$ and ε the set of partial index-derivatives is $\emptyset$. Let ρ be a_i. We need to prove that $\forall \mathrm{I} \in \mathcal{I}_{a_i} \forall x \in \mathcal{I}_{a_i}^\star \ (\partial_{x\mathrm{I}}(a_i) \neq \emptyset \implies \partial_{x\mathrm{I}}(a_i) = \{\mathsf{c}_{\mathrm{I}}(a_i)\} \neq \{\emptyset\})$. Let $\partial_{x\mathrm{I}}(a_i) \neq \emptyset$, then by Proposition 26, $\partial_{x\mathrm{I}}(a_i) = \{\varepsilon\}$ and $x\mathrm{I} = \{i\}$. Then $\mathrm{I} = \{i\}$ and $\mathsf{c}_{\mathrm{I}}(a_i) = \varepsilon$. Thus, we conclude that $\partial_{x\mathrm{I}}(a_i) = \{\mathsf{c}_{\mathrm{I}}(a_i)\} \neq \{\emptyset\}$.

Let us suppose that for ρ_i, $i = 1, 2$ we have $\forall \mathrm{I} \in \mathcal{I}_{\rho_i} \forall x \in \mathcal{I}_{\rho_i}^\star \ (\partial_{x\mathrm{I}}(\rho_i) \neq \emptyset \implies \partial_{x\mathrm{I}}(\rho_i) = \{\mathsf{c}_{\mathrm{I}}(\rho_i)\} \neq \{\emptyset\})$.

Let $\rho = \rho_1 + \rho_2$ be such that $\partial_{x\mathrm{I}}(\rho_1 + \rho_2) \neq \emptyset$. Then, $\partial_{x\mathrm{I}}(\rho_1 + \rho_2) = \partial_{x\mathrm{I}}(\rho_i)$ with $x\mathrm{I} = (x\mathrm{I})\big|_{\rho_i}$, for some $i \in \{1, 2\}$. By the induction hypothesis, $\partial_{x\mathrm{I}}(\rho_i) = \{\mathsf{c}_{\mathrm{I}}(\rho_i)\} \neq \{\emptyset\}$. Thus, $\mathsf{c}_{\mathrm{I}}(\rho_i) \neq \emptyset$ and $\mathsf{c}_{\mathrm{I}}(\rho_1 + \rho_2) = \mathsf{c}_{\mathrm{I}}(\rho_i)$.

Let $\rho = \rho_1\rho_2$. If $\partial_{x\mathrm{I}}(\rho_1\rho_2) \neq \emptyset$ then we have to consider two cases. Let $\partial_{x\mathrm{I}}(\rho_1\rho_2) = \partial_{x\mathrm{I}}(\rho_1) \odot \rho_2$ and $x\mathrm{I} = (x\mathrm{I})\big|_{\rho_1}$. Then, $\partial_{x\mathrm{I}}(\rho_1) \neq \emptyset$ and $\partial_{x\mathrm{I}}(\rho_1) = \{\mathsf{c}_{\mathrm{I}}(\rho_1)\}$. We conclude that $\mathsf{c}_{\mathrm{I}}(\rho_1) \neq \emptyset$ and $\mathsf{c}_{\mathrm{I}}(\rho_1\rho_2) = \mathsf{c}_{\mathrm{I}}(\rho_1)$. In the second case, $\partial_{x\mathrm{I}}(\rho_1\rho_2) = \partial_{z\mathrm{I}}(\rho_2) \neq \emptyset$, $x = yz$, $\varepsilon(\partial_y(\rho_1)) = \varepsilon$ and $z\mathrm{I} = (z\mathrm{I})\big|_{\rho_2}$. We conclude that $y = y\big|_{\rho_1}$ and $\mathrm{I} = \mathrm{I}\big|_{\rho_2}$. Then, $\mathsf{c}_{\mathrm{I}}(\rho_1) = \emptyset$ and $\mathsf{c}_{\mathrm{I}}(\rho_1\rho_2) = \mathsf{c}_{\mathrm{I}}(\rho_2)$. By the induction hypothesis, $\partial_{z\mathrm{I}}(\rho_2) = \{\mathsf{c}_{\mathrm{I}}(\rho_2)\}$ and the result follows.

Let $\rho = \rho_1^\star$. If $\partial_{x\mathrm{I}}(\rho_1^\star) \neq \emptyset$, we can write $\partial_{x\mathrm{I}}(\rho_1^\star) = \partial_{v_1\mathrm{I}}(\rho_1) \odot \rho_1^\star \cup \cdots \cup \partial_{v_n\mathrm{I}}(\rho_1) \odot \rho_1^\star$, with $n \geq 1$, such that for all $1 \leq i \leq n$, $x = u_iv_i$ and $\partial_{v_i\mathrm{I}}(\rho_1) \odot \rho_1^\star \neq \emptyset$. By the induction hypothesis, each nonempty set of partial index-derivatives $\partial_{v_i\mathrm{I}}(\rho_1)$ is equal to $\{\mathsf{c}_{\mathrm{I}}(\rho_1)\} \neq \{\emptyset\}$. Thus, $\partial_{x\mathrm{I}}(\rho_1^\star) = \{\mathsf{c}_{\mathrm{I}}(\rho_1)\rho_1^\star\}$.

Finally, let $\rho = \rho_1 \cap \rho_2$ be such that $\partial_{x\mathrm{I}}(\rho_1 \cap \rho_2) \neq \emptyset$. Then $\partial_{x\mathrm{I}}(\rho_1 \cap \rho_2) = \partial_{(x\mathrm{I})|_{\rho_1}}(\rho_1) \Cap \partial_{(x\mathrm{I})|_{\rho_2}}(\rho_2)$, $x\mathrm{I} = (x\mathrm{I})\big|_{\rho_1} \cap_{\mathcal{I}} (x\mathrm{I})\big|_{\rho_2}$ and $\partial_{(x\mathrm{I})|_{\rho_i}}(\rho_i) \neq \emptyset$, for $i = 1, 2$. Moreover, $\partial_{(x\mathrm{I})|_{\rho_i}}(\rho_i) = \{\mathsf{c}_{\mathrm{I}|_{\rho_i}}(\rho_i)\}$. The result follows by the induction hypothesis and from the definition of $\mathsf{c}_{\mathrm{I}}(\rho_1 \cap \rho_2)$. ☐

This result guarantees that, given a linear indexed expression ρ and $\mathrm{I} \in \mathcal{I}_\rho$, all sets of partial index-derivatives $\partial_{x\mathrm{I}}(\rho)$ different from $\emptyset$ are singletons with an unique c-continuation $\mathsf{c}_{\mathrm{I}}(\rho)$ of ρ w.r.t. I.

The next lemmata justify the construction of the c-continuation automaton (Definition 31) which will be proved to be isomorphic to $\mathcal{A}_{\mathsf{posi}}$.

Lemma 29. *Consider a linear indexed expression ρ. Then, $\mathrm{I} \in \mathsf{Lst}(\rho)$ if and only if $\varepsilon(\mathsf{c}_{\mathrm{I}}(\rho)) = \varepsilon$.*

Proof. Since $\mathsf{c}_{\mathrm{I}}(\rho) \neq \emptyset$ is a consequence of $\varepsilon(\mathsf{c}_{\mathrm{I}}(\rho)) = \varepsilon$, it is sufficient to prove the equivalence of $\mathrm{I} \in \mathsf{Lst}(\rho)$ and $\varepsilon(\mathsf{c}_{\mathrm{I}}(\rho)) = \varepsilon$ by structural induction on ρ.

For $\rho = a_i$ and $\mathrm{I} = \{i\} \in \mathsf{Lst}(\rho)$, we have $\varepsilon(\mathsf{c}_{\mathrm{I}}(\rho)) = \varepsilon(\varepsilon) = \varepsilon$.

For $\rho = \rho_1 + \rho_2$, we have

$$\begin{aligned}\mathrm{I} \in \mathsf{Lst}(\rho) &\iff \mathrm{I} \in \mathsf{Lst}(\rho_1) \ \vee \ \mathrm{I} \in \mathsf{Lst}(\rho_2)\\ &\iff \varepsilon(\mathsf{c}_\mathrm{I}(\rho_1)) = \varepsilon \ \text{ or } \ \varepsilon(\mathsf{c}_\mathrm{I}(\rho_2)) = \varepsilon\\ &\iff \varepsilon(\mathsf{c}_\mathrm{I}(\rho)) = \varepsilon.\end{aligned}$$

Now, consider $\rho = \rho_1 \cdot \rho_2$. One has

$$\begin{aligned}\mathrm{I} \in \mathsf{Lst}(\rho) &\iff (\mathrm{I} \in \mathsf{Lst}(\rho_1) \wedge \varepsilon(\rho_2) = \varepsilon) \vee \mathrm{I} \in \mathsf{Lst}(\rho_2)\\ &\iff (\varepsilon(\mathsf{c}_\mathrm{I}(\rho_1)) = \varepsilon \wedge \varepsilon(\rho_2) = \varepsilon) \vee \varepsilon(\mathsf{c}_\mathrm{I}(\rho_2)) = \varepsilon\\ &\iff \varepsilon(\mathsf{c}_\mathrm{I}(\rho_1) \cdot \rho_2) = \varepsilon \vee \varepsilon(\mathsf{c}_\mathrm{I}(\rho_2)) = \varepsilon\\ &\iff \varepsilon(\mathsf{c}_\mathrm{I}(\rho)) = \varepsilon.\end{aligned}$$

The case for $\rho = \rho_1^\star$ is straightforward. Finally, for $\rho = \rho_1 \cap \rho_2$, we have

$$\begin{aligned}\mathrm{I} \in \mathsf{Lst}(\rho_1 \cap \rho_2) &\iff \mathrm{I}\big|_{\rho_1} \in \mathsf{Lst}(\rho_1) \wedge \mathrm{I}\big|_{\rho_2} \in \mathsf{Lst}(\rho_2) \wedge \mathrm{I} = \mathrm{I}\big|_{\rho_1} \cup \mathrm{I}\big|_{\rho_2}\\ &\iff \varepsilon(\mathsf{c}_{\mathrm{I}|_{\rho_1}}(\rho_1)) = \varepsilon \wedge \varepsilon(\mathsf{c}_{\mathrm{I}|_{\rho_2}}(\rho_2)) = \varepsilon \wedge \mathrm{I} = \mathrm{I}\big|_{\rho_1} \cup \mathrm{I}\big|_{\rho_2}\\ &\iff \varepsilon(\mathsf{c}_\mathrm{I}(\rho)) = \varepsilon.\end{aligned}$$

□

Lemma 30. *Consider a linear indexed expression ρ and sets of indexes $\mathrm{I}, \mathrm{J} \in \mathcal{I}_\rho$. Then, $(\mathrm{I}, \mathrm{J}) \in \mathsf{Fol}(\rho)$ if and only if $\mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho))$.*

Proof. Throughout the proof, by structural induction on ρ, we will use the fact that for $\rho = \rho_1 + \rho_2$, as well as for $\rho = \rho_1 \cdot \rho_2$, at most one of $\mathrm{I} \subseteq \mathsf{ind}(\rho_i)$ $(i = 1, 2)$ is true, and consequently at most one of $\mathsf{c}_\mathrm{I}(\rho_i) \neq \emptyset$ holds. Also note that $\mathrm{J} \in \mathsf{Fst}(\tau)$ implies that $\tau \neq \emptyset$.

Let $\rho = \rho_1 + \rho_2$. Then,

$$\begin{aligned}(\mathrm{I}, \mathrm{J}) \in \mathsf{Fol}(\rho) &\iff (\mathrm{I}, \mathrm{J}) \in \mathsf{Fol}(\rho_1) \vee (\mathrm{I}, \mathrm{J}) \in \mathsf{Fol}(\rho_2)\\ &\iff \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_1)) \vee \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_2)) \iff \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho)).\end{aligned}$$

Now, consider $\rho = \rho_1 \cdot \rho_2$. One has

$$\begin{aligned}(\mathrm{I}, \mathrm{J}) \in \mathsf{Fol}(\rho) &\iff (\mathrm{I} \in \mathsf{Lst}(\rho_1) \wedge \mathrm{J} \in \mathsf{Fst}(\rho_2)) \vee (\mathrm{I}, \mathrm{J}) \in \mathsf{Fol}(\rho_1) \vee (\mathrm{I}, \mathrm{J}) \in \mathsf{Fol}(\rho_2)\\ &\iff \begin{aligned}&(\varepsilon(\mathsf{c}_\mathrm{I}(\rho_1)) = \varepsilon \wedge \mathrm{J} \in \mathsf{Fst}(\rho_2)) \vee \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_1)) \vee\\ &\qquad \vee \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_2))\end{aligned}\\ &\iff \begin{aligned}&(\varepsilon(\mathsf{c}_\mathrm{I}(\rho_1)) = \varepsilon \wedge \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_1)) \cup \mathsf{Fst}(\rho_2)) \ \vee \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_2)) \vee\\ &\qquad \vee \ (\mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_1)) \wedge \mathsf{c}_\mathrm{I}(\rho_1) \neq \emptyset)\end{aligned}\\ &\iff (\mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_1) \cdot \rho_2) \wedge \mathsf{c}_\mathrm{I}(\rho_1) \neq \emptyset) \ \vee \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_2))\\ &\iff \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho)),\end{aligned}$$

where the third equivalence results from distributing the disjunct $\mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_1))$ over the first conjunction. For the fourth equivalence the definitions of Fst and c_I were used.

Now, consider $\rho = \rho_1^\star$. Then,

$$\begin{aligned}(\mathrm{I},\mathrm{J}) \in \mathsf{Fol}(\rho) &\iff (\mathrm{I},\mathrm{J}) \in \mathsf{Fol}(\rho_1) \vee \ (\mathrm{I} \in \mathsf{Lst}(\rho_1) \wedge \ \mathrm{J} \in \mathsf{Fst}(\rho_1))\\ &\iff \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_1)) \vee (\varepsilon(\mathsf{c}_\mathrm{I}(\rho_1)) = \varepsilon \wedge \mathrm{J} \in \mathsf{Fst}(\rho_1))\\ &\iff \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_1)\rho_1^\star) \iff \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho)).\end{aligned}$$

Finally, let $\rho = \rho_1 \cap \rho_2$. Note that for every union $\mathrm{I} = \mathrm{I}_1 \cup \mathrm{I}_2$, the condition $\ell(\mathrm{I}_1) = \ell(\mathrm{I}_2)$ is true by definition, since $\mathrm{I} \in \mathcal{I}_\rho$. The same holds for J. Then,

$$\begin{aligned}(\mathrm{I},\mathrm{J}) \in \mathsf{Fol}(\rho) &\iff \begin{array}{l}\exists \mathrm{I}_1,\mathrm{J}_1 \in \mathcal{I}_{\rho_1} \exists \mathrm{I}_2,\mathrm{J}_2 \in \mathcal{I}_{\rho_2}\,((\mathrm{I}_1,\mathrm{J}_1) \in \mathsf{Fol}(\rho_1) \wedge \\ \qquad \wedge\, (\mathrm{I}_2,\mathrm{J}_2) \in \mathsf{Fol}(\rho_2) \wedge \mathrm{I} = \mathrm{I}_1 \cup \mathrm{I}_2 \wedge \mathrm{J} = \mathrm{J}_1 \cup \mathrm{J}_2)\end{array}\\ &\iff \begin{array}{l}\exists \mathrm{I}_1,\mathrm{J}_1 \in \mathcal{I}_{\rho_1} \exists \mathrm{I}_2,\mathrm{J}_2 \in \mathcal{I}_{\rho_2}(\mathrm{J}_1 \in \mathsf{Fst}(\mathsf{c}_{\mathrm{I}_1}(\rho_1)) \wedge \\ \qquad \wedge\, \mathrm{J}_2 \in \mathsf{Fst}(\mathsf{c}_{\mathrm{I}_2}(\rho_2)) \wedge \mathrm{I} = \mathrm{I}_1 \cup \mathrm{I}_2 \wedge \mathrm{J} = \mathrm{J}_1 \cup \mathrm{J}_2)\end{array}\\ &\iff \begin{array}{l}\exists \mathrm{J}_1 \in \mathcal{I}_{\rho_1} \exists \mathrm{J}_2 \in \mathcal{I}_{\rho_2}(\mathrm{J}_1 \in \mathsf{Fst}(\mathsf{c}_{\mathrm{I}|_{\rho_1}}(\rho_1)) \wedge \\ \qquad \wedge\, \mathrm{J}_2 \in \mathsf{Fst}(\mathsf{c}_{\mathrm{I}|_{\rho_2}}(\rho_2)) \wedge \mathrm{J} = \mathrm{J}_1 \cup \mathrm{J}_2 \wedge \mathrm{I} = \mathrm{I}\big|_{\rho_1} \cup \mathrm{I}\big|_{\rho_1})\end{array}\\ &\iff \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_{\mathrm{I}|_{\rho_1}}(\rho_1) \cap \mathsf{c}_{\mathrm{I}|_{\rho_2}}(\rho_2)) \wedge \mathrm{I} = \mathrm{I}\big|_{\rho_1} \cup \mathrm{I}\big|_{\rho_2}\\ &\iff \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho)).\end{aligned}$$

□

Definition 31. The c-continuation automaton of an expression $\alpha \in \mathsf{RE}_\cap$ is

$$\mathcal{A}_\mathsf{c}(\alpha) = \langle S_\mathsf{c}, \Sigma, \{(\{0\}, \mathsf{c}_{\{0\}}(\overline{\alpha}))\}, \delta_\mathsf{c}, F_\mathsf{c}\rangle,$$

where $S_\mathsf{c} = \{\ (\mathrm{I}, \mathsf{c}_\mathrm{I}(\overline{\alpha})) \mid \mathrm{I} \in S_\mathsf{posi}\ \}$, $F_\mathsf{c} = \{\ (\mathrm{I}, \mathsf{c}_\mathrm{I}(\overline{\alpha})) \mid \varepsilon(\mathsf{c}_\mathrm{I}(\overline{\alpha})) = \varepsilon\ \}$, $\mathsf{c}_{\{0\}}(\overline{\alpha}) = \overline{\alpha}$, $\delta_\mathsf{c} = \{\ ((\mathrm{I}, \mathsf{c}_\mathrm{I}(\overline{\alpha})), \ell(\mathrm{J}), (\mathrm{J}, \mathsf{c}_\mathrm{J}(\overline{\alpha}))) \mid \mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\overline{\alpha}))\ \}$.

By Lemma 29, Lemma 30, and considering $\varphi : S_\mathsf{c} \to S_\mathsf{posi}$ such that $\varphi((\mathrm{I}, \mathsf{c}_\mathrm{I}(\overline{\alpha}))) = \mathrm{I}$, the following holds.

Theorem 32. *For $\alpha \in \mathsf{RE}_\cap$, we have $\mathcal{A}_\mathsf{posi}(\alpha) \simeq \mathcal{A}_\mathsf{c}(\alpha)$.*

Example 33. Consider the expression $\overline{\alpha} = (b_1 a_2^\star b_3 + a_4) \cap (a_5 a_6 + b_7)^\star$, from Example 18, and let $\rho_2 = (a_5 a_6 + b_7)^\star$. We have the following c-continuations: $\mathsf{c}_{\{1,7\}}(\overline{\alpha}) = a_2^\star b_3 \cap \rho_2$, $\mathsf{c}_{\{4,5\}}(\overline{\alpha}) = \varepsilon \cap a_6\rho_2$, $\mathsf{c}_{\{4,6\}}(\overline{\alpha}) = \varepsilon \cap \rho_2$, $\mathsf{c}_{\{2,5\}}(\overline{\alpha}) = a_2^\star b_3 \cap a_6\rho_2$, $\mathsf{c}_{\{2,6\}}(\overline{\alpha}) = a_2^\star b_3 \cap \rho_2$, and $\mathsf{c}_{\{3,7\}}(\overline{\alpha}) = \varepsilon \cap \rho_2$.

6. $\mathcal{A}_\mathsf{pd}$ as a Quotient of $\mathcal{A}_\mathsf{pos}$

Using $\mathcal{A}_\mathsf{c}$ we show that the partial derivative automaton $\mathcal{A}_\mathsf{pd}$ is a quotient of $\mathcal{A}_\mathsf{pos}$. This extends the corresponding result for simple regular expressions, although the proof cannot use the same technique. Recall that, for a simple regular expression α, one builds $\mathcal{A}_\mathsf{pd}(\overline{\alpha})$, and then shows that when its transitions are unmarked, the result $\overline{\mathcal{A}_\mathsf{pd}(\overline{\alpha})}$ is isomorphic to a quotient of $\mathcal{A}_\mathsf{c}(\alpha)$. However, with $\alpha \in \mathsf{RE}_\cap$, this method cannot be used because, as mentioned in the introduction, intersection does not commute with marking. For $\alpha \in \mathsf{RE}_\cap$, we will present a direct isomorphism between $\mathcal{A}_{\mathsf{pd}}(\alpha)$ and a quotient of $\mathcal{A}_{\mathsf{c}}(\alpha)$. The next lemmas will be needed to build that isomorphism.

Lemma 34. *Consider a linear indexed expression ρ and $\mathrm{I} \in \mathcal{I}_\rho$. If $\mathrm{I} \in \mathsf{Fst}(\rho)$, then $\mathsf{c}_\mathrm{I}(\rho) \neq \emptyset$ and $\mathsf{c}_\mathrm{I}(\rho) \in \partial_\mathrm{I}(\rho)$.*

Proof. We proceed by structural induction on ρ. For $\rho \in \{\varepsilon, \emptyset\}$ there is nothing to prove. For $\rho = a_i$ the statement is obviously true. Suppose that the statement is true for ρ_1 and ρ_2. Let $\rho = \rho_1 + \rho_2$. If $\mathrm{I} \in \mathsf{Fst}(\rho_i)$, $\emptyset \neq \mathsf{c}_\mathrm{I}(\rho) = \mathsf{c}_\mathrm{I}(\rho_i) \in \partial_\mathrm{I}(\rho_i) \subseteq \partial_\mathrm{I}(\rho)$, for either $i = 1$ or $i = 2$. Let $\rho = \rho_1 \cdot \rho_2$. If $\mathrm{I} \in \mathsf{Fst}(\rho_1)$, then $\mathsf{c}_\mathrm{I}(\rho_1) \neq \emptyset$ and $\mathsf{c}_\mathrm{I}(\rho) = \mathsf{c}_\mathrm{I}(\rho_1) \cdot \rho_2 \in \partial_\mathrm{I}(\rho_1) \odot \rho_2 \subseteq \partial_\mathrm{I}(\rho)$. If $\mathrm{I} \in \mathsf{Fst}(\rho_2)$, then $\varepsilon(\rho_1) = \varepsilon$ and $\mathsf{c}_\mathrm{I}(\rho_1) = \emptyset$. We have $\emptyset \neq \mathsf{c}_\mathrm{I}(\rho) = \mathsf{c}_\mathrm{I}(\rho_2) \in \partial_\mathrm{I}(\rho_2) \subseteq \partial_\mathrm{I}(\rho)$. For $\rho = \rho_1^\star$ the result is immediate. Finally, let $\rho = \rho_1 \cap \rho_2$. If $\mathrm{I} \in \mathsf{Fst}(\rho_1 \cap \rho_2)$, then $\mathrm{I} = \mathrm{I}_1 \cup \mathrm{I}_2$ with $\mathrm{I}_1 \in \mathsf{Fst}(\rho_1)$ and $\mathrm{I}_2 \in \mathsf{Fst}(\rho_2)$. The result easily follows from the induction hypothesis and the definitions. ☐

Lemma 35. *Consider a linear indexed expression ρ and $\mathrm{I}, \mathrm{J} \in \mathcal{I}_\rho$, such that $\mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho))$. Then, $\mathsf{c}_\mathrm{J}(\rho) \in \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho))$.*

Proof. We proceed by structural induction on ρ. There is nothing to prove for $\rho \in \{\emptyset, \varepsilon\} \cup \Sigma_\rho$.

Consider $\rho = \rho_1 + \rho_2$ and let $\mathsf{c}_\mathrm{I}(\rho) = \mathsf{c}_\mathrm{I}(\rho_i)$, for either $i = 1$ or $i = 2$. Since $\mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho)) = \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_i))$, we conclude that $\mathrm{J} \subseteq \mathsf{ind}(\mathsf{c}_\mathrm{I}(\rho_i)) \subseteq \mathsf{ind}(\rho_i)$ and consequently $\mathsf{c}_\mathrm{J}(\rho) = \mathsf{c}_\mathrm{J}(\rho_i)$. By the induction hypothesis, $\mathsf{c}_\mathrm{J}(\rho_i) \in \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho_i))$, i.e. $\mathsf{c}_\mathrm{J}(\rho) \in \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho))$.

Now, let $\rho = \rho_1 \cdot \rho_2$. For $\mathsf{c}_\mathrm{I}(\rho_1 \cdot \rho_2) = \mathsf{c}_\mathrm{I}(\rho_1) \cdot \rho_2$ and $\mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_1) \cdot \rho_2)$, there are two cases to consider. First, if $\mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_1))$, then $\mathrm{J} \subseteq \mathsf{ind}(\rho_1)$ and, by the induction hypothesis, $\mathsf{c}_\mathrm{J}(\rho_1) \in \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho_1))$. Thus, $\mathsf{c}_\mathrm{J}(\rho) = \mathsf{c}_\mathrm{J}(\rho_1) \cdot \rho_2 \in \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho_1)) \odot \rho_2 \subseteq \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho_1) \cdot \rho_2) = \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho))$. On the other hand, if $\mathrm{J} \in \mathsf{Fst}(\rho_2)$, then $\varepsilon(\mathsf{c}_\mathrm{I}(\rho_1)) = \varepsilon$ and also $\mathrm{J} \subseteq \mathsf{ind}(\rho_2)$. Then, $\mathsf{c}_\mathrm{J}(\rho) = \mathsf{c}_\mathrm{J}(\rho_2) \in \partial_\mathrm{J}(\rho_2) \subseteq \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho_1) \cdot \rho_2) = \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho))$. Finally, if $\mathsf{c}_\mathrm{I}(\rho) = \mathsf{c}_\mathrm{I}(\rho_2)$ and $\mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_2))$, we have $\mathsf{c}_\mathrm{J}(\rho) = \mathsf{c}_\mathrm{J}(\rho_2) \in \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho_2)) = \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho))$.

For $\rho = \rho_1^\star$ and $\mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho_1))$, we have $\mathsf{c}_\mathrm{J}(\rho_1) \in \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho_1))$ and consequently $\mathsf{c}_\mathrm{J}(\rho) = \mathsf{c}_\mathrm{J}(\rho_1) \cdot \rho \in \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho_1)) \odot \rho \subseteq \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho_1) \cdot \rho) = \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\rho))$. On the other hand, if $\mathrm{J} \in \mathsf{Fst}(\rho) = \mathsf{Fst}(\rho_1)$, then $\varepsilon(\mathsf{c}_\mathrm{I}(\rho_1)) = \varepsilon$. Thus, $\mathsf{c}_\mathrm{J}(\rho_1) \in \partial_\mathrm{J}(\rho_1)$ and $\mathsf{c}_\mathrm{J}(\rho) = \mathsf{c}_\mathrm{J}(\rho_1) \cdot \rho \in \partial_\mathrm{J}(\rho_1) \odot \rho = \partial_\mathrm{J}(\rho)$.

Finally, let $\rho = \rho_1 \cap \rho_2$ and $\mathsf{c}_\mathrm{I}(\rho) = \mathsf{c}_{\mathrm{I}_1}(\rho_1) \cap \mathsf{c}_{\mathrm{I}_2}(\rho_2)$, where $\mathrm{I} = \mathrm{I}_1 \cup \mathrm{I}_2$ and $\mathrm{I}_i = \mathrm{I}\big|_{\rho_i}$ for $i = 1, 2$. Since $\mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\rho)) = \mathsf{Fst}(\mathsf{c}_{\mathrm{I}_1}(\rho_1)) \cup \mathsf{Fst}(\mathsf{c}_{\mathrm{I}_2}(\rho_2))$, we have $\mathrm{J} = \mathrm{J}_1 \cup \mathrm{J}_2$, where $\mathrm{J}_i \in \mathsf{Fst}(\mathsf{c}_{\mathrm{I}_i}(\rho_i))$ and $\mathrm{J}_i \in \mathsf{ind}(\mathsf{c}_{\mathrm{I}_i}(\rho_i)) \subseteq \mathsf{ind}(\rho_i)$, for $i = 1, 2$. By the induction hypothesis $\mathsf{c}_{\mathrm{J}_i}(\rho_i) \in \partial_{\mathrm{J}_i}(\mathsf{c}_{\mathrm{I}_i}(\rho_i))$, thus $\mathsf{c}_\mathrm{J}(\rho) = \mathsf{c}_{\mathrm{J}_1}(\rho_1) \cap \mathsf{c}_{\mathrm{J}_2}(\rho_2) \in \partial_{\mathrm{J}_1}(\rho_1) \Cap \partial_{\mathrm{J}_2}(\rho_2) = \partial(\rho)$. ☐

Lemma 36. *Consider well-indexed expressions ρ', ρ and $\mathrm{I} \in \mathcal{I}_\rho$, such that $\rho' \in \partial_\mathrm{I}(\rho)$. Then, $\overline{\rho'} \in \partial_{\ell(\mathrm{I})}(\overline{\rho})$.*

Proof. The proof procedes by induction on the structure of ρ. We only present the case for $\rho = \rho_1 \cap \rho_2$. If $\rho_1' \cap \rho_2' \in \partial_\mathrm{I}(\rho_1 \cap \rho_2)$, then $\overline{\rho_i'} \in \partial_{\ell(\mathrm{I})}(\overline{\rho_i})$ for $i = 1, 2$. Thus,

$\overline{\rho_1' \cap \rho_2'} = \overline{\rho_1'} \cap \overline{\rho_2'} \in \partial_{\ell(\mathrm{I})}(\overline{\rho_1}) \Cap \partial_{\ell(\mathrm{I})}(\overline{\rho_2}) = \partial_{\ell(\mathrm{I})}(\overline{\rho_1 \cap \rho_2})$. $\square$

Lemma 37. *Consider a well-indexed expression ρ, $a \in \Sigma$ and $\beta \in \partial_a(\overline{\rho})$. Then, there exist $\mathrm{I} \in \mathcal{I}_\rho$ and $\rho' \in \partial_\mathrm{I}(\rho)$ with $\ell(\mathrm{I}) = a$ and $\overline{\rho'} = \beta$. Furthermore, for $x = a_1 \cdots a_n \in \Sigma^\star$, if $\beta \in \partial_x(\overline{\rho})$, there exist $\mathrm{I}_1 \cdots \mathrm{I}_n \in \mathcal{I}_\rho^\star$ and $\rho' \in \partial_{\mathrm{I}_1 \cdots \mathrm{I}_n}(\rho)$ with $\ell(\mathrm{I}_1 \cdots \mathrm{I}_n) = x$ and $\overline{\rho'} = \beta$.*

Proof. The proof is straightforward by induction on the structure of ρ. We present only the case for $\rho = \rho_1 \cap \rho_2$. Let $\beta_1 \cap \beta_2 \in \partial_a(\overline{\rho_1 \cap \rho_2}) = \partial_a(\overline{\rho_1}) \Cap \partial_a(\overline{\rho_2})$, for some letter a. It follows from the induction hypothesis that there are $\mathrm{I}_i \in \mathcal{I}_{\rho_i}$ and $\rho_i' \in \partial_{\mathrm{I}_i}(\rho_i)$, with $\ell(\mathrm{I}_i) = a$, for $i = 1, 2$, and such that $\overline{\rho_1'} = \beta_1$ and $\overline{\rho_2'} = \beta_2$. Thus, $\mathrm{I} = \mathrm{I}_1 \cup \mathrm{I}_2 \in \mathcal{I}_\rho$, $\mathrm{I}\big|_{\rho_i} = \mathrm{I}_i$ (ρ is well-indexed), $\ell(\mathrm{I}) = a$, $\rho_1' \cap \rho_2' \in \partial_\mathrm{I}(\rho_1 \cap \rho_2)$ and $\overline{\rho_1' \cap \rho_2'} = \beta_1 \cap \beta_2$. $\square$

Given $\alpha \in \mathsf{RE}_\cap$, consider $\mathcal{A}_\mathsf{c}(\alpha)$ and the equivalence relation $\equiv_\ell$ on S_c given by $(\mathrm{I}, \mathsf{c}_\mathrm{I}(\overline{\alpha})) \equiv_\ell (\mathrm{J}, \mathsf{c}_\mathrm{J}(\overline{\alpha}))$ if and only if $\overline{\mathsf{c}_\mathrm{I}(\overline{\alpha})} = \overline{\mathsf{c}_\mathrm{J}(\overline{\alpha})}$, for $\mathrm{I}, \mathrm{J} \in \mathcal{I}_{\overline{\alpha}} \cup \{\{0\}\}$.

Lemma 38. *The relation $\equiv_\ell$ is right invariant w.r.t. $\mathcal{A}_\mathsf{c}$.*

Proof. Let $(\mathrm{I}, \mathsf{c}_\mathrm{I}(\overline{\alpha})) \equiv_\ell (\mathrm{J}, \mathsf{c}_\mathrm{J}(\overline{\alpha}))$. First, note that $\varepsilon(\mathsf{c}_\mathrm{I}(\overline{\alpha})) = \varepsilon$ if and only if $\varepsilon(\overline{\mathsf{c}_\mathrm{I}(\overline{\alpha})}) = \varepsilon$. Now, let $(\mathrm{I}_1, \mathsf{c}_{\mathrm{I}_1}(\overline{\alpha})) \in \delta_\mathsf{c}((\mathrm{I}, \mathsf{c}_\mathrm{I}(\overline{\alpha})), \ell(\mathrm{I}_1))$ with $\mathrm{I}_1 \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\overline{\alpha}))$. By Lemma 35, we have $\mathsf{c}_{\mathrm{I}_1}(\overline{\alpha}) \in \partial_{\mathrm{I}_1}(\mathsf{c}_\mathrm{I}(\overline{\alpha}))$, and by Lemma 36, $\overline{\mathsf{c}_{\mathrm{I}_1}(\overline{\alpha})} \in \partial_{\ell(\mathrm{I}_1)}(\overline{\mathsf{c}_\mathrm{I}(\overline{\alpha})}) = \partial_{\ell(\mathrm{I}_1)}(\overline{\mathsf{c}_\mathrm{J}(\overline{\alpha})})$. Then, there exist $\mathrm{I}_2 \in \mathcal{I}_{\mathsf{c}_\mathrm{J}(\overline{\alpha})}$ and $\rho' \in \partial_{\mathrm{I}_2}(\mathsf{c}_\mathrm{J}(\overline{\alpha})) \subseteq \partial_{\mathrm{I}_2}(\partial_\mathrm{J}(\overline{\alpha})) = \partial_{\mathrm{J}\mathrm{I}_2}(\overline{\alpha})$ with $\overline{\rho'} = \overline{\mathsf{c}_{\mathrm{I}_1}(\overline{\alpha})}$ by Lemma 37. But, by Proposition 28, $\rho' = \mathsf{c}_{\mathrm{I}_2}(\overline{\alpha})$. $\square$

Theorem 39. *For $\alpha \in \mathsf{RE}_\cap$,*

$$\mathcal{A}_\mathsf{pd}(\alpha) \simeq \mathcal{A}_\mathsf{c}(\alpha)^\mathsf{ac}/{\equiv_\ell}.$$

Proof. Let $\mathcal{A}_\mathsf{c}(\alpha)^\mathsf{ac}/{\equiv_\ell} = (S_\ell, \Sigma, \delta_\ell, [(\{0\}, \overline{\alpha})], F_\ell)$. Consider the map

$$\begin{aligned} \varphi : \quad S_\ell \quad &\longrightarrow \quad \partial(\alpha) \\ [(\mathrm{I}, \mathsf{c}_\mathrm{I}(\overline{\alpha}))] &\longmapsto \overline{\mathsf{c}_\mathrm{I}(\overline{\alpha})}. \end{aligned}$$

To show that φ is an isomorphism we need that:

1) φ is well-defined;
2) φ is bijective;
3) $\varphi(\delta_\ell(s, a)) = \delta_\mathsf{pd}(\varphi(s), a)$ for every $s \in S_\ell, a \in \Sigma$;
4) $\varphi(F_\ell) = F_\mathsf{pd}$;
5) $\varphi([(\{0\}, \mathsf{c}_{\{0\}}(\overline{\alpha}))]) = \alpha$.

Claim 1) follows from lemmas 34 and 35. For 3) we consider both inclusions. Consider $\beta \in \varphi(\delta_\ell(s, a))$, for $s \in S_\ell$ and $a \in \Sigma$. Then, there exist $\mathrm{I}, \mathrm{J} \in \mathcal{I}_{\overline{\alpha}}$ such that $[(\mathrm{I}, \mathsf{c}_\mathrm{I}(\overline{\alpha}))] = s$, $\overline{\mathsf{c}_\mathrm{J}(\overline{\alpha})} = \beta$, $(\mathrm{J}, \mathsf{c}_\mathrm{J}(\overline{\alpha})) \in \delta_\mathsf{c}((\mathrm{I}, \mathsf{c}_\mathrm{I}(\overline{\alpha})), \ell(\mathrm{J}))$ and $\ell(\mathrm{J}) = a$, i.e.

$\mathrm{J} \in \mathsf{Fst}(\mathsf{c}_\mathrm{I}(\overline{\alpha}))$. By Lemma 35, we have $\mathsf{c}_\mathrm{J}(\overline{\alpha}) \in \partial_\mathrm{J}(\mathsf{c}_\mathrm{I}(\overline{\alpha}))$ and by Lemma 36, $\overline{\mathsf{c}_\mathrm{J}(\overline{\alpha})} \in \partial_a(\overline{\mathsf{c}_\mathrm{I}(\overline{\alpha})})$. Thus, $\overline{\mathsf{c}_\mathrm{J}(\overline{\alpha})} \in \delta_{\mathsf{pd}}(\overline{\mathsf{c}_\mathrm{I}(\overline{\alpha})}, a)$. Now, let $\beta \in \delta_{\mathsf{pd}}(\tau, a)$, where $\tau = \overline{\mathsf{c}_\mathrm{I}(\overline{\alpha})}$, for some $\mathrm{I} \in \mathcal{I}_{\overline{\alpha}}$ and $a \in \Sigma$. Then, there is a sequence of terms $\tau_0 = \alpha, \tau_1, \ldots, \tau_n = \tau$ and a sequence of letters $a_1, \ldots, a_n \in \Sigma$ such that $\tau_{i+1} \in \partial_{a_{i+1}}(\tau_i)$, for $0 \leq i \leq n-1$, and $\beta \in \partial_a(\tau)$, i.e. $\beta \in \partial_{a_1 \cdots a_n a}(\alpha)$. By Lemma 37, there exist $\mathrm{J}_1, \ldots, \mathrm{J}_n, \mathrm{J} \in \mathcal{I}_{\overline{\alpha}}$, with $\ell(\mathrm{J}_1 \cdots \mathrm{J}_n \mathrm{J}) = a_1 \cdots a_n a$, and $\rho' \in \partial_{\mathrm{J}_1 \cdots \mathrm{J}_n \mathrm{J}}(\overline{\alpha})$ such that $\overline{\rho'} = \beta$. By Proposition 28, $\rho' = \mathsf{c}_\mathrm{J}(\overline{\alpha})$. On the other hand, it is straightforward to show by induction on the structure of a well-indexed expression ρ, that $\partial_\mathrm{J}(\rho) \neq \emptyset$ implies $\mathrm{J} \in \mathsf{Fst}(\rho)$. Thus, $[(\mathrm{J}, \mathsf{c}_\mathrm{J}(\overline{\alpha}))] \in \delta_\ell([(\mathrm{I}, \mathsf{c}_\mathrm{I}(\overline{\alpha}))], \ell(\mathrm{J}))$ and consequently $\beta = \overline{\mathsf{c}_\mathrm{J}(\overline{\alpha})} \in \varphi(\delta_\ell([(\mathrm{I}, \mathsf{c}_\mathrm{I}(\overline{\alpha}))], a))$. □

Example 40. Consider $\alpha = (ba^\star b + a) \cap (aa + b)^\star$ from examples 18, 22, and 33. Set $\beta = (aa + b)^\star$. For the positions present in $\mathcal{A}_\mathsf{c}(\alpha)^{\mathsf{ac}}$, we have $\overline{\mathsf{c}_{\{4,5\}}(\overline{\alpha})} = \varepsilon \cap a\beta$, $\overline{\mathsf{c}_{\{3,7\}}(\overline{\alpha})} = \varepsilon \cap \beta$, $\overline{\mathsf{c}_{\{2,5\}}(\overline{\alpha})} = a^\star b \cap a\beta$, and $\overline{\mathsf{c}_{\{1,7\}}(\overline{\alpha})} = \overline{\mathsf{c}_{\{2,6\}}(\overline{\alpha})} = a^\star b \cap \beta$. Merging states $(\{1,7\}, \mathsf{c}_{\{1,7\}}(\overline{\alpha}))$ and $(\{2,6\}, \mathsf{c}_{\{2,6\}}(\overline{\alpha}))$ in $\mathcal{A}_\mathsf{c}(\alpha)^{\mathsf{ac}}$, one obtains an NFA isomorphic to $\mathcal{A}_{\mathsf{pd}}(\alpha)$, which is represented in Figure 2.

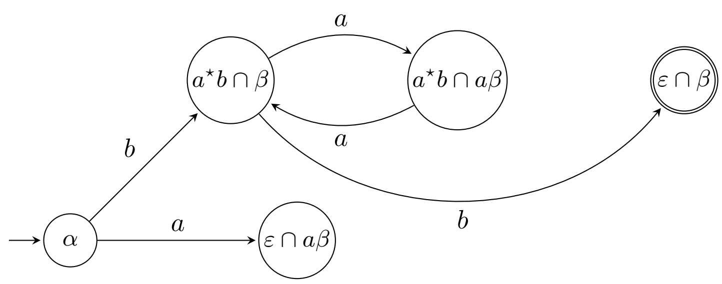

Figure 2: $\mathcal{A}_{\mathsf{pd}}((ba^\star b + a) \cap (aa + b)^\star)$

7. Final Remarks

For simple regular expressions of size n, the size of $\mathcal{A}_{\mathsf{pos}}(\alpha)$ is $O(n^2)$, and using $\mathcal{A}_\mathsf{c}(\alpha)$ it is possible to efficiently compute $\mathcal{A}_{\mathsf{pd}}(\alpha)$ [11]. For regular expressions with intersection the conversion to NFA's has exponential computational complexity [13] and both the size of $\mathcal{A}_{\mathsf{pos}}$ and $\mathcal{A}_{\mathsf{pd}}$ can be exponential in the size of the regular expression. On the average case, however, the size of these automata seem to be much smaller [3], and thus feasible for practical applications. In this scenario, algorithms for building $\mathcal{A}_{\mathsf{pd}}$ using $\mathcal{A}_{\mathsf{pos}}$ seem worthwhile to develop.

References

[1] V. Antimirov, Partial Derivatives of Regular Expressions and Finite Automaton Constructions. *Theoret. Comput. Sci.* **155** (1996) 2, 291–319.

[2] R. Bastos, S. Broda, A. Machiavelo, N. Moreira, R. Reis, On the State Complexity of Partial Derivative Automaton for Regular Expressions with Intersection. In: C. Câmpeanu, F. Manea, J. Shallit (eds.), *18th DCFS*. LNCS 9777, Springer, 2016, 45–59.

[3] R. Bastos, S. Broda, A. Machiavelo, N. Moreira, R. Reis, On the Average Complexity of Partial Derivative Automata for Semi-extended Expressions. *Journal of Automata, Languages and Combinatorics* **22** (2017) 1-3, 5–28.

[4] G. Berry, R. Sethi, From Regular Expressions to Deterministic Automata. *Theoret. Comput. Sci.* **48** (1986), 117–126.

[5] S. Broda, A. Machiavelo, N. Moreira, , R. Reis, Position Automaton Construction for Regular Expressions with Intersection. In: S. Brlek, C. Reutenauer (eds.), *20th DLT*. LNCS 9840, Springer, 2016, 51–63.

[6] A. Brüggemann-Klein, Regular Expressions into Finite Automata. *Theoret. Comput. Sci.* **48** (1993), 197–213.

[7] J. Brzozowski, Derivatives of Regular Expressions. *JACM* **11** (1964) 4, 481–494.

[8] P. Caron, J. Champarnaud, L. Mignot, Partial Derivatives of an Extended Regular Expression. In: A. H. Dediu, S. Inenaga, C. Martín-Vide (eds.), *5th LATA*. LNCS 6638, Springer, 2011, 179–191.

[9] P. Caron, J. Champarnaud, L. Mignot, A General Framework for the Derivation of Regular Expressions. *RAIRO - Theor. Inf. and Applic.* **48** (2014) 3, 281–305.

[10] P. Caron, D. Ziadi, Characterization of Glushkov Automata. *Theoret. Comput. Sci.* **233** (2000) 1-2, 75–90.

[11] J. M. Champarnaud, D. Ziadi, Canonical Derivatives, Partial Derivatives and Finite Automaton Constructions. *Theoret. Comput. Sci.* **289** (2002), 137–163.

[12] H. Chen, S. Yu, Derivatives of Regular Expressions and an Application. In: M. J. Dinneen, B. Khoussainov, A. Nies (eds.), *Computation, Physics and Beyond, WTCS 2012*. LNCS 7160, Springer, 2012, 343–356.

[13] W. Gelade, Succinctness of Regular Expressions with Interleaving, Intersection and Counting. *Theor. Comput. Sci.* **411** (2010) 31-33, 2987–2998.

[14] V. M. Glushkov, The Abstract Theory of Automata. *Russian Math. Surveys* **16** (1961), 1–53.

[15] J. E. Hopcroft, J. D. Ullman, *Introduction to Automata Theory, Languages and Computation*. Addison Wesley, 1979.

[16] L. Ilie, S. Yu, Follow Automata. *Inf. Comput.* **186** (2003) 1, 140–162.

[17] R. McNaughton, H. Yamada, Regular Expressions and State Graphs for Automata. *IEEE Transactions on Electronic Computers* **9** (1960), 39–47.

[18] J. Sakarovitch, *Elements of Automata Theory*. Cambridge University Press, 2009.

[19] S. Yu, Regular languages. In: G. Rozenberg, A. Salomaa (eds.), *Handbook of Formal Languages*. 1, Springer, 1997, 41–110.