Average State Complexity of Partial Derivative Automata for Synchronised Shuffles

Sabine Broda, António Machiavelo, Nelma Moreira, and Rogério Reis *

CMUP & DM-DCC, Faculdade de Ciências da Universidade do Porto
Rua do Campo Alegre, 4169-007 Porto, Portugal
{sabine.broda,antonio.machiavelo,nelma.moreira,rogerio.reis}@fc.up.pt

Synchronised shuffle operators allow to specify symbols on which the operands must or can synchronise instead of interleave. Recently, partial derivative and position based automata for regular expressions with synchronised shuffle operators were introduced. In this paper, using the framework of analytic combinatorics, we study the asymptotic average state complexity of partial derivative automata for regular expressions with strongly and arbitrarily synchronised shuffles. The new results extend and improve the ones previously obtained for regular expressions with shuffle and intersection, as these operations can be seen as special cases of synchronised shuffles. For intersection, asymptotically the average state complexity of the partial derivative automaton is 3, which significantly improves the known exponential upper-bound.

1. Introduction

Synchronised shuffle operators allow to specify symbols on which the operands must or can synchronise instead of interleave. Intersection and shuffle can be seen as two extreme cases, corresponding to strict synchronisation and pure interleaving. Several variants were introduced and studied by ter Beek et al. [3], motivated by modelling synchronisation in concurrent systems or certain gene operations in molecular biology. Sulzmann and Thiemann [15] studied regular expressions with a general synchronised shuffling operator and extended the notions of derivatives and partial derivatives to these expressions. Broda et al. [8] defined a location based position automaton for regular expressions with strongly, arbitrarily, and weakly synchronised operators and showed that the partial derivative automaton (defined in [15]) is a quotient of the position automaton. For a standard regular expression α, the partial derivative automaton $\mathcal{A}_{\mathrm{PD}}(\alpha)$, introduced by Antimirov [1], can also be obtained by solving a system of equations, whose solution is a support set $\pi(\alpha)$, due to Mirkin [13]. In fact, the set of states of $\mathcal{A}_{\mathrm{PD}}(\alpha)$ is equal to $\pi(\alpha) \cup \{\alpha\}$. In Bastos et al. [2], the rules for computing $\pi(\alpha)$ were extended to regular expressions with intersection. However, in that case it was shown that both constructions were not identical, as Mirkin's construction could have some states that were not accessible. Still, the inductive definition of a support set is essential to obtain average

*This work was partially supported by CMUP, which is financed by national funds through FCT under the project with reference UIDB/00144/2020.

complexity results using the framework of analytic combinatorics [4, 2, 5, 6]. In the worst-case, the number of states of a partial derivative automaton for expressions with shuffle or intersection can be exponential in the size of the expression. In the average-case, the known upper bounds are smaller, but still exponential [2, 5]. In this paper, we define new rules for computing the support for a system of expression equations, where for each alphabet symbol the solution is independently computed. In the case of standard regular expressions, the support obtained by these rules is the same as the one defined by Mirkin. However, the new rules allow to consider synchronising and non-synchronising symbols separately, and thereby obtain a support for a regular expression with strongly and arbitrarily synchronised operators. Moreover, the new rules lead to a smaller support for the intersection operator [2], which corresponds to the strong synchronised operator when all alphabet symbols synchronise. Using the framework of analytic combinatorics, we give an upper bound for the asymptotic size of the support. We also estimate the asymptotic average number of partial derivatives by one symbol. In particular, for intersection we show that asymptotically, as the size of the alphabet grows, the size of the support set, thus the average state complexity of $\mathcal{A}_{\mathrm{PD}}$, is 3. This is a surprising improvement with regard to the previous upper bound of $(1.056+o(1))^n$, where n is the size of the expression [2]. However, experimental results in [2], suggested that the size of $\mathcal{A}_{\mathrm{PD}}$ as the alphabet size grows could approach the constant 3. This paper extends [7] by presenting full proofs of all our results, by obtaining an asymptotic estimate for the average size of the support for the arbitrarily synchronised shuffle, and by comparing the average complexity results obtained for the intersection operator with experimental ones.

2. Regular Expressions with Synchronised Shuffles

Let $\Sigma = \{\sigma_1, \ldots, \sigma_m\}$ be an alphabet and $\Sigma^\star$ be the set of words over Σ. A *language* is any subset of $\Sigma^\star$. The *empty word* is denoted by ε. The set of alphabet symbols that occur in a word $w \in \Sigma^\star$ is denoted by Σ_w. Given a set $\Gamma \subseteq \Sigma$, the *strongly synchronised shuffle* of two words w.r.t. Γ imposes synchronisation on all symbols of Γ. Formally, the *strongly synchronised shuffle* of two words for $u, v \in \Sigma^\star$, w.r.t. Γ, and denoted by $u \,{}^{\mathsf{s}}\|_\Gamma\, v$ is defined inductively as follows [15]:

$$\varepsilon \,{}^{\mathsf{s}}\|_\Gamma\, v = v \,{}^{\mathsf{s}}\|_\Gamma\, \varepsilon = \begin{cases} \{v\}, & \text{if } \Sigma_v \cap \Gamma = \emptyset, \\ \emptyset, & \text{otherwise,} \end{cases}$$

$$\sigma u \,{}^{\mathsf{s}}\|_\Gamma\, \tau v = \begin{cases} \{\, \sigma w \mid w \in u \,{}^{\mathsf{s}}\|_\Gamma\, v \,\}, & \text{if } \sigma = \tau \wedge \sigma \in \Gamma, \\ \emptyset, & \text{if } \sigma \neq \tau \wedge \sigma, \tau \in \Gamma, \\ \{\, \sigma w \mid w \in u \,{}^{\mathsf{s}}\|_\Gamma\, \tau v \,\}, & \text{if } \sigma \notin \Gamma \wedge \tau \in \Gamma, \\ \{\, \tau w \mid w \in \sigma u \,{}^{\mathsf{s}}\|_\Gamma\, v \,\}, & \text{if } \sigma \in \Gamma \wedge \tau \notin \Gamma, \\ \{\, \sigma w \mid w \in u \,{}^{\mathsf{s}}\|_\Gamma\, \tau v \,\} & \\ \quad \cup \{\, \tau w \mid w \in \sigma u \,{}^{\mathsf{s}}\|_\Gamma\, v \,\}, & \text{if } \sigma, \tau \notin \Gamma. \end{cases}$$

For $\Gamma = \emptyset$ the operator ${}^{\mathsf{s}}\|_{\emptyset}$ coincides with the usual shuffle operator $\sqcup\!\sqcup$, given by $u \sqcup\!\sqcup \varepsilon = \varepsilon \sqcup\!\sqcup u = \{u\}$ and $\sigma u \sqcup\!\sqcup \tau v = \{\, \sigma w \mid w \in u \sqcup\!\sqcup \tau v \,\} \cup \{\, \tau w \mid w \in \sigma u \sqcup\!\sqcup v \,\}$, for $u, v \in \Sigma^\star$ and $\sigma, \tau \in \Sigma$. On the other hand, if $\Gamma = \Sigma$, the operator ${}^{\mathsf{s}}\|_{\Sigma}$ corresponds to intersection ($\cap$). The *arbitrarily synchronised shuffle* of two words w.r.t. $\Gamma \subseteq \Sigma$ permits symbols in Γ to synchronise, but does not force their synchronisation. Formally, the *arbitrarily synchronised shuffle* of words for $u, v \in \Sigma^\star$, denoted by $u \,{}^{\mathsf{a}}\|_{\Gamma}\, v$, is defined as follows [3]:

$$
\begin{aligned}
&\varepsilon \,{}^{\mathsf{a}}\|_{\Gamma}\; v = v \,{}^{\mathsf{a}}\|_{\Gamma}\, \varepsilon = \{v\},\\
&\sigma u \,{}^{\mathsf{a}}\|_{\Gamma}\, \tau v = \begin{cases} \{\, \sigma w \mid w \in u \,{}^{\mathsf{a}}\|_{\Gamma}\, \tau v \,\} \cup \{\, \tau w \mid w \in \sigma u \,{}^{\mathsf{a}}\|_{\Gamma}\, v \,\}, & \text{if } \sigma \neq \tau \vee \sigma \notin \Gamma,\\ \{\, \sigma w \mid w \in u \,{}^{\mathsf{a}}\|_{\Gamma}\, \tau v \,\} \cup \{\, \tau w \mid w \in \sigma u \,{}^{\mathsf{a}}\|_{\Gamma}\, v \,\} & \\ \quad \cup \{\, \sigma w \mid w \in u \,{}^{\mathsf{a}}\|_{\Gamma}\, v \,\}, & \text{if } \sigma = \tau \wedge \sigma \in \Gamma. \end{cases}
\end{aligned}
$$

Example 1. *We have* $abca \,{}^{\mathsf{s}}\|_{\{a\}}\, ada = \{abcda, abdca, adbca\}$, $ab \,{}^{\mathsf{s}}\|_{\{a\}}\, da = \{dab\}$, *and* $ab \,{}^{\mathsf{a}}\|_{\{a\}}\, da = \{abda, adba, adab, dab, daab, daba\}$.

Given two languages $L_1, L_2 \subseteq \Sigma^\star$ and $\circ \in \{\, {}^{\mathsf{s}}\|_{\Gamma}\,, {}^{\mathsf{a}}\|_{\Gamma} \,\}$ one has, as usual, $L_1 \circ L_2 = \bigcup_{u \in L_1, v \in L_2} u \circ v$. If L_1 and L_2 are regular, $L_1 \circ L_2$ is regular [3]. Both ${}^{\mathsf{s}}\|_{\Gamma}$ and ${}^{\mathsf{a}}\|_{\Gamma}$ are commutative and associative, and distribute over $\cup$. The set of regular expressions with synchronised shuffles over the alphabet Σ, RE($\|$), contains $\emptyset$ and is generated by the following grammar

$$
\alpha \to \varepsilon \mid \sigma \in \Sigma \mid (\alpha + \alpha) \mid (\alpha\alpha) \mid (\alpha^\star) \mid (\alpha \,{}^{\mathsf{s}}\|_{\Gamma}\, \alpha) \mid (\alpha \,{}^{\mathsf{a}}\|_{\Gamma}\, \alpha). \tag{1}
$$

Let RE be the subset of RE($\|$) of standard regular expressions without the operators ${}^{\mathsf{s}}\|_{\Gamma}$ and ${}^{\mathsf{a}}\|_{\Gamma}$. The *language* associated with an expression $\alpha \in$ RE($\|$) is denoted by $\mathcal{L}(\alpha)$, which for $\alpha \in$ RE is defined as usual, and $\mathcal{L}(\alpha_1 \circ \alpha_2) = \mathcal{L}(\alpha_1) \circ \mathcal{L}(\alpha_2)$ for $\circ \in \{\, {}^{\mathsf{s}}\|_{\Gamma}\,, {}^{\mathsf{a}}\|_{\Gamma} \,\}$. Two regular expressions α and β are *equivalent*, and we write $\alpha \doteq \beta$, iff $\mathcal{L}(\alpha) = \mathcal{L}(\beta)$. We define $\varepsilon(\alpha)$ by $\varepsilon(\alpha) = \varepsilon$ if $\varepsilon \in \mathcal{L}(\alpha)$, and $\varepsilon(\alpha) = \emptyset$, otherwise. In the same way, given a language L one defines $\varepsilon(L)$. Given a set of expressions S, the *language* associated with S is $\mathcal{L}(S) = \bigcup_{\alpha \in S} \mathcal{L}(\alpha)$. If $\beta \in$ RE($\|$) $\setminus \{\varepsilon, \emptyset\}$, we define $S\beta = \{\alpha\beta \mid \alpha \in S \wedge \alpha \neq \varepsilon\} \cup (\varepsilon \in S)\{\beta\}$. We have $\varepsilon S = S\varepsilon = S$ and $\emptyset S = S\emptyset = \emptyset$. Moreover, for $S, T \subseteq$ RE($\|$) $\setminus \{\emptyset\}$ and $\circ \in \{\, {}^{\mathsf{s}}\|_{\Gamma}\,, {}^{\mathsf{a}}\|_{\Gamma} \,\}$, we define $S \circ T = \{\, \alpha \circ \beta \mid \alpha \in S \wedge \beta \in T \,\}$. The size of $\alpha \in$ RE($\|$) is denoted by $|\alpha|$ and defined as the number of occurrences of symbols (parenthesis not counted) in α.

3. Automata and Systems of Equations

A *nondeterministic finite automaton* (NFA) is a quintuple $A = \langle Q, \Sigma, \delta, I, F \rangle$ where Q is a finite set of states, Σ is a finite alphabet, $I \subseteq Q$ is the set of initial states, $F \subseteq Q$ is the set of final states, and $\delta : Q \times \Sigma \to 2^Q$ is the transition function. The function δ can be naturally extended to sets of states and to words. In what follows we will take $Q = [1, n]$, for $|Q| = n$. The *language* of A is $\mathcal{L}(A) = \{\, w \in \Sigma^\star \mid \delta(I, w) \cap F \neq \emptyset \,\}$. The *right language* of a state q, denoted by $\mathcal{L}_q$, is the language accepted by A if we take $I = \{q\}$. It is well known that it is possible to associate to each n-state NFA A over $\Sigma = \{\sigma_1, \dots, \sigma_m\}$, with right languages $\mathcal{L}_1, \dots, \mathcal{L}_n$, a system of linear language equations

$$\mathcal{L}_i = \sigma_1 \mathcal{L}_{1i} \cup \dots \cup \sigma_m \mathcal{L}_{mi} \cup \varepsilon(\mathcal{L}_i), \quad \text{for } i \in Q,$$

where $\mathcal{L}_{ji} = \bigcup_{h \in \delta(i,\sigma_j)} \mathcal{L}_h$ and $\mathcal{L}(A) = \bigcup_{i \in I} \mathcal{L}_i$. In the same way, it is possible to associate to each regular expression a system of equations. Given an expression α over Σ, a *support* for $\alpha = \alpha_0$ is a set of expressions $\{\alpha_1, \dots, \alpha_n\}$ that satisfies a system of equations

$$\alpha_i \doteq \sigma_1 \alpha_{i,1} + \dots + \sigma_m \alpha_{i,m} + \varepsilon(\alpha_i), \quad i \in [0, n] \tag{2}$$

where each of $\alpha_{i,1}, \dots, \alpha_{i,m}$ is a (possibly empty) sum of elements in $\{\alpha_1, \dots, \alpha_n\}$. Mirkin [13, 9] showed that for a standard regular expression $\alpha \in \mathrm{RE}$, a support $\pi(\alpha)$ can be computed as follows:

$$\begin{aligned} \pi(\emptyset) &= \pi(\varepsilon) = \emptyset, & \pi(\alpha + \beta) &= \pi(\alpha) \cup \pi(\beta), \\ \pi(\sigma) &= \{\varepsilon\} \quad (\sigma \in \Sigma), & \pi(\alpha\beta) &= \pi(\alpha)\beta \cup \pi(\beta), \\ \pi(\alpha^\star) &= \pi(\alpha)\alpha^\star. \end{aligned} \tag{3}$$

In the following, we show that the support of an expression α, $\pi(\alpha)$, can be written as the union of m sets, each set corresponding to a letter $\sigma_j \in \Sigma$ ($j \in [1, m]$) and denoted by $\pi^{\sigma_j}(\alpha)$, i.e., $\pi(\alpha) = \bigcup_{\sigma_j \in \Sigma} \pi^{\sigma_j}(\alpha)$. Moreover, each set $\pi^{\sigma_j}(\alpha)$ can be obtained independently, and contains exactly the components in the sums $\alpha_{0,j}, \dots, \alpha_{n,j}$ in (2). For standard regular expressions this result is established in Lemma 2.

Lemma 2. *Consider a standard regular expression α over an alphabet Σ and $\pi(\alpha) = \{\alpha_1, \dots, \alpha_n\}$ its support. Then, we have $\pi(\alpha) = \bigcup_{\sigma \in \Sigma} \pi^\sigma(\alpha)$, where:*

$$\begin{aligned} \pi^\sigma(\emptyset) &= \pi^\sigma(\varepsilon) = \emptyset, & \pi^\sigma(\alpha + \beta) &= \pi^\sigma(\alpha) \cup \pi^\sigma(\beta), \\ \pi^\sigma(\sigma) &= \{\varepsilon\}, & \pi^\sigma(\alpha\beta) &= \pi^\sigma(\alpha)\beta \cup \pi^\sigma(\beta), \\ \pi^\sigma(\tau) &= \emptyset \quad (\tau \neq \sigma). & \pi^\sigma(\alpha^\star) &= \pi^\sigma(\alpha)\alpha^\star. \end{aligned}$$

Furthermore, $\pi(\alpha)$ satisfies a system of equations of the form,

$$\alpha_i \doteq \sigma_1 \alpha_{i,1} + \dots + \sigma_m \alpha_{i,m} + \varepsilon(\alpha_i), \quad i \in [0, n] \tag{4}$$

such that each $\alpha_{i,j}$ is a (possibly empty) sum of elements in $\pi^{\sigma_j}(\alpha)$, for $j \in [1, m]$.

Proof. Using (3) we prove the result by induction on the structure of the regular expression. For ε and $\emptyset$, it is obvious. For $\sigma \in \Sigma$, we have $\pi(\sigma) = \pi^\sigma(\sigma) = \{\varepsilon\}$ and $\sigma \doteq \sigma\varepsilon$. Now, suppose that the result holds for α_0 and β_0, and consider $\pi(\alpha_0) = \{\alpha_1, \dots, \alpha_{n_1}\}$ and $\pi(\beta_0) = \{\beta_1, \dots, \beta_{n_2}\}$ satisfying respectively the set of equations

$$\alpha_i \doteq \sigma_1 \alpha_{i,1} + \dots + \sigma_m \alpha_{i,m} + \varepsilon(\alpha_i), \quad i \in [0, n_1]$$

and

$$\beta_j \doteq \sigma_1 \beta_{j,1} + \dots + \sigma_m \beta_{j,m} + \varepsilon(\beta_j), \quad j \in [0, n_2]$$

where, for $s = 1, \dots, m$ we have that $\alpha_{i,s}$ and $\beta_{j,s}$ are linear combinations of elements of $\pi^{\sigma_s}(\alpha_0)$ and of $\pi^{\sigma_s}(\beta_0)$, respectively.

For $\alpha_0 + \beta_0$ we have

$$\pi(\alpha_0 + \beta_0) = \pi(\alpha_0) \cup \pi(\beta_0) = \bigcup_{\sigma \in \Sigma} \pi^{\sigma}(\alpha_0) \ \cup \ \bigcup_{\sigma \in \Sigma} \pi^{\sigma}(\beta_0) = \bigcup_{\sigma \in \Sigma} \pi^{\sigma}(\alpha_0 + \beta_0).$$

It remains to show that the elements of $\{\alpha_0 + \beta_0\} \cup \pi(\alpha_0 + \beta_0) = \{\alpha_0 + \beta_0\} \cup \pi(\alpha_0) \cup \pi(\beta_0)$ satisfy a set of equations as in (4). For $\alpha_i \in \pi(\alpha_0)$ this follows from the induction hypothesis and since $\pi^{\sigma}(\alpha_0) \subseteq \pi^{\sigma}(\alpha_0 + \beta_0)$, for $\sigma \in \Sigma$. Analogously, for $\beta_j \in \pi(\beta_0)$. Finally, $\alpha_0 + \beta_0$ satisfies

$$\alpha_0 + \beta_0 \doteq \sigma_1(\alpha_{0,1} + \beta_{0,1}) + \cdots + \sigma_m(\alpha_{0,m} + \beta_{0,m}) + \varepsilon(\alpha_0 + \beta_0),$$

where $\alpha_{0,s} + \beta_{0,s}$ are linear combinations of elements of $\pi^{\sigma_s}(\alpha_0 + \beta_0) = \pi^{\sigma_s}(\alpha_0) \cup \pi^{\sigma_s}(\beta_0)$ for $s \in [1, m]$.

For $\alpha_0\beta_0$ we have

$$\begin{aligned}\pi(\alpha_0\beta_0) = \pi(\alpha_0)\beta_0 \cup \pi(\beta_0) &= \bigcup_{\sigma \in \Sigma} \pi^{\sigma}(\alpha_0)\beta_0 \cup \bigcup_{\sigma \in \Sigma} \pi^{\sigma}(\beta_0) \\ &= \bigcup_{\sigma \in \Sigma} (\pi^{\sigma}(\alpha_0)\beta_0 \cup \pi^{\sigma}(\beta_0)) = \bigcup_{\sigma \in \Sigma} \pi^{\sigma}(\alpha_0\beta_0).\end{aligned}$$

Furthermore, expression $\alpha_0\beta_0$ satisfies

$$\begin{aligned}\alpha_0\beta_0 &\doteq \sigma_1(\alpha_{0,1}\beta_0) + \cdots + \sigma_m(\alpha_{0,m}\beta_0) + \varepsilon(\alpha_0)\beta_0 \\ &\doteq \sigma_1(\alpha_{0,1}\beta_0 + \varepsilon(\alpha_0)\beta_{0,1}) + \cdots + \sigma_m(\alpha_{0,m}\beta_0 + \varepsilon(\alpha_0)\beta_{0,m}) + \varepsilon(\alpha_0\beta_0),\end{aligned}$$

where, using distributivity of $\cdot$ over $+$, $\alpha_{0,s}\beta_0 + \varepsilon(\alpha_0)\beta_{0,s}$ are equivalent to linear combinations of elements of $\pi^{\sigma_s}(\alpha_0\beta_0) = \pi^{\sigma_s}(\alpha_0)\beta_0 \cup \pi^{\sigma_s}(\beta_0)$ for $s \in [1, m]$. The case of elements in $\pi(\alpha_0)\beta_0$ is shown analogously, and the case of elements of $\pi(\beta_0)$ follows by induction.

For $\alpha_0^\star$ we have

$$\pi(\alpha_0^\star) = \pi(\alpha_0)\alpha_0^\star = \left(\bigcup_{\sigma \in \Sigma} \pi^{\sigma}(\alpha_0)\right)\alpha_0^\star = \bigcup_{\sigma \in \Sigma} \pi^{\sigma}(\alpha_0)\alpha_0^\star = \bigcup_{\sigma \in \Sigma} \pi^{\sigma}(\alpha_0^\star).$$

Furthermore, expression $\alpha_0^\star$ satisfies

$$\begin{aligned}\alpha_0^\star &\doteq (\sigma_1\alpha_{0,1} + \cdots + \sigma_m\alpha_{0,m} + \varepsilon(\alpha_0))^\star \\ &\doteq (\sigma_1\alpha_{0,1} + \cdots + \sigma_m\alpha_{0,m})^\star \doteq (\sigma_1\alpha_{0,1} + \cdots + \sigma_m\alpha_{0,m})\alpha_0^\star + \varepsilon \\ &\doteq \sigma_1(\alpha_{0,1}\alpha_0^\star) + \cdots + \sigma_m(\alpha_{0,m}\alpha_0^\star) + \varepsilon(\alpha_0^\star),\end{aligned}$$

where, using distributivity as above, $\alpha_{0,s}\alpha_0^\star$ are equivalent to linear combinations of $\pi^{\sigma_s}(\alpha_0^\star) = \pi^{\sigma_s}(\alpha_0)\alpha_0^\star$, for $s \in [1, m]$. The case of elements in $\pi(\alpha_0)\alpha_0^\star$ is shown analogously. □

We extend the notion of support to $\alpha \in \mathrm{RE}(\|)$, which in the case of ${}^{\mathsf{s}}\|_{\Sigma}$, i.e. of intersection, represents a significant improvement with regard to the definition of a support given in [2]. For $\circ \in \{\, {}^{\mathsf{s}}\|_{\Gamma}\,,\, {}^{\mathsf{a}}\|_{\Gamma}\,\}$ we define

$$\pi(\alpha \circ \beta) = \bigcup_{\sigma \in \Sigma} \pi^{\sigma}(\alpha \circ \beta), \tag{5}$$

where

$$\pi^{\sigma}(\alpha \,{}^{\mathsf{s}}\|_{\Gamma}\, \beta) = \begin{cases} \pi^{\sigma}(\alpha) \,{}^{\mathsf{s}}\|_{\Gamma}\, \pi^{\sigma}(\beta), & \text{for } \sigma \in \Gamma; \\ \pi^{\sigma}(\alpha) \,{}^{\mathsf{s}}\|_{\Gamma}\, (\pi(\beta) \cup \{\beta\}) \cup (\pi(\alpha) \cup \{\alpha\}) \,{}^{\mathsf{s}}\|_{\Gamma}\, \pi^{\sigma}(\beta), & \text{otherwise;} \end{cases} \tag{6}$$

$$\pi^{\sigma}(\alpha \,{}^{\mathsf{a}}\|_{\Gamma}\, \beta) = \begin{cases} \pi^{\sigma}(\alpha) \,{}^{\mathsf{a}}\|_{\Gamma}\, (\pi(\beta) \cup \{\beta\}) \;\cup\; (\pi(\alpha) \cup \{\alpha\}) \,{}^{\mathsf{a}}\|_{\Gamma}\, \pi^{\sigma}(\beta) & \\ \qquad\qquad \cup\, \pi^{\sigma}(\alpha) \,{}^{\mathsf{a}}\|_{\Gamma}\, \pi^{\sigma}(\beta), & \text{for } \sigma \in \Gamma; \\ \pi^{\sigma}(\alpha) \,{}^{\mathsf{a}}\|_{\Gamma}\, (\pi(\beta) \cup \{\beta\}) \;\cup\; (\pi(\alpha) \cup \{\alpha\}) \,{}^{\mathsf{a}}\|_{\Gamma}\, \pi^{\sigma}(\beta), & \text{otherwise.} \end{cases} \tag{7}$$

Using (5) one obtains a support for $\alpha \in \mathrm{RE}(\|)$.

Proposition 3. *Given $\alpha \in \mathrm{RE}(\|)$, $\pi(\alpha) = \{\alpha_1, \ldots, \alpha_n\}$ is a support for $\alpha = \alpha_0$, satisfying a system of equations of the form,*

$$\alpha_i \doteq \sigma_1 \alpha_{i,1} + \cdots + \sigma_m \alpha_{i,m} + \varepsilon(\alpha_i), \quad i \in [0, n] \tag{8}$$

such that each $\alpha_{i,s}$ is a (possibly empty) sum of elements in $\pi^{\sigma_s}(\alpha)$, for $s \in [1, m]$.

Proof. Considering Lemma 2 and its proof, it suffices to prove equation (8) for an expression of the form $\alpha_0 \circ \beta_0$, with $\circ \in \{\, {}^{\mathsf{s}}\|_{\Gamma}\,,\, {}^{\mathsf{a}}\|_{\Gamma}\,\}$, supposing that the result holds for α_0 and for β_0. As such, let $\pi(\alpha_0) = \{\alpha_1, \ldots, \alpha_{n_1}\}$ and $\pi(\beta_0) = \{\beta_1, \ldots, \beta_{n_2}\}$ be a support for α_0 and for β_0, respectively, satisfying the following equations.

$$\alpha_i \doteq \sigma_1 \alpha_{i,1} + \cdots + \sigma_m \alpha_{i,m} + \varepsilon(\alpha_i), \quad i \in [0, n_1],$$

and

$$\beta_j \doteq \sigma_1 \beta_{j,1} + \cdots + \sigma_m \beta_{j,m} + \varepsilon(\beta_j), \quad j \in [0, n_2],$$

where, for $s \in [1, m]$ we have that $\alpha_{i,s}$ and $\beta_{j,s}$ are linear combinations of elements of $\pi^{\sigma_s}(\alpha_0)$ and of $\pi^{\sigma_s}(\beta_0)$, respectively. Without loss of generality we suppose that $m = \ell + k$, and that $\sigma_1, \ldots, \sigma_\ell \in \Gamma$, and $\sigma_{\ell+1}, \ldots, \sigma_{\ell+k} \in \Sigma \setminus \Gamma$.

We start by considering the case $\circ = {}^{\mathsf{s}}\|_{\Gamma}$. Note, that every member of $\pi(\alpha_0 \,{}^{\mathsf{s}}\|_{\Gamma}\, \beta_0) \cup \{\alpha_0 \,{}^{\mathsf{s}}\|_{\Gamma}\, \beta_0\}$ is of the form $\alpha_i \,{}^{\mathsf{s}}\|_{\Gamma}\, \beta_j$, where $i \in [0, n_1]$ and $j \in [0, n_2]$. We have the following:

$$\begin{aligned}
\alpha_i \,{}^{\mathsf{s}}\|_\Gamma\, \beta_j \doteq\; & (\sigma_1\alpha_{i,1} + \cdots + \sigma_\ell\alpha_{i,\ell} + \sigma_{\ell+1}\alpha_{i,\ell+1} + \cdots + \sigma_{\ell+k}\alpha_{i,\ell+k} + \varepsilon(\alpha_i)) \\
& {}^{\mathsf{s}}\|_\Gamma\, (\sigma_1\beta_{j,1} + \cdots + \sigma_\ell\beta_{j,\ell} + \sigma_{\ell+1}\beta_{j,\ell+1} + \cdots + \sigma_{\ell+k}\beta_{j,\ell+k} + \varepsilon(\beta_j)) \\
& \doteq \sigma_1(\alpha_{i,1} \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,1}) + \cdots + \sigma_\ell(\alpha_{i,\ell} \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,\ell}) \\
& + \sigma_{\ell+1}(\alpha_{i,\ell+1} \,{}^{\mathsf{s}}\|_\Gamma\, \beta_j) + \cdots + \sigma_{\ell+k}(\alpha_{i,\ell+k} \,{}^{\mathsf{s}}\|_\Gamma\, \beta_j) \\
& + \sigma_{\ell+1}(\alpha_i \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,\ell+1}) + \cdots + \sigma_{\ell+k}(\alpha_i \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,\ell+k}) \\
& + \varepsilon(\alpha_i) \,{}^{\mathsf{s}}\|_\Gamma\, \sigma_1\beta_{j,1} + \cdots + \varepsilon(\alpha_i) \,{}^{\mathsf{s}}\|_\Gamma\, \sigma_{\ell+k}\beta_{j,\ell+k} \\
& + \sigma_1\alpha_{i,1} \,{}^{\mathsf{s}}\|_\Gamma\, \varepsilon(\beta_j) + \cdots + \sigma_{\ell+k}\alpha_{i,\ell+k} \,{}^{\mathsf{s}}\|_\Gamma\, \varepsilon(\beta_j) + \varepsilon(\alpha_i) \,{}^{\mathsf{s}}\|_\Gamma\, \varepsilon(\beta_j) \\
\doteq\; & \sigma_1(\alpha_{i,1} \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,1}) + \cdots + \sigma_\ell(\alpha_{i,\ell} \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,\ell}) + \sigma_{\ell+1}(\alpha_{i,\ell+1} \,{}^{\mathsf{s}}\|_\Gamma\, \beta_j + \alpha_i \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,\ell+1}) \\
& + \cdots + \sigma_{\ell+k}(\alpha_{i,\ell+k} \,{}^{\mathsf{s}}\|_\Gamma\, \beta_j + \alpha_i \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,\ell+k}) + \varepsilon(\alpha_i \,{}^{\mathsf{s}}\|_\Gamma\, \beta_j).
\end{aligned}$$

The last equality follows on one hand from $\varepsilon(\alpha_i) \,{}^{\mathsf{s}}\|_\Gamma\, \varepsilon(\beta_j) \doteq \varepsilon(\alpha_i \,{}^{\mathsf{s}}\|_\Gamma\, \beta_j)$. On the other, we have

$$\begin{gathered}
\varepsilon(\alpha_i) \,{}^{\mathsf{s}}\|_\Gamma\, \sigma_1\beta_{j,1} + \cdots + \varepsilon(\alpha_i) \,{}^{\mathsf{s}}\|_\Gamma\, \sigma_{\ell+k}\beta_{j,\ell+k} \doteq \\
\varepsilon(\alpha_i) \,{}^{\mathsf{s}}\|_\Gamma\, \sigma_{\ell+1}\beta_{j,\ell+1} + \cdots + \varepsilon(\alpha_i) \,{}^{\mathsf{s}}\|_\Gamma\, \sigma_{\ell+k}\beta_{j,\ell+k} \doteq \\
\sigma_{\ell+1}(\varepsilon(\alpha_i) \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,\ell+1}) + \cdots + \sigma_{\ell+k}(\varepsilon(\alpha_i) \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,\ell+k}),
\end{gathered}$$

and furthermore $\mathcal{L}(\varepsilon(\alpha_i) \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,\ell+1}) \subseteq \mathcal{L}(\alpha_i \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,\ell+1})$. In the similar way we simplify $\sigma_1\alpha_{i,1} \,{}^{\mathsf{s}}\|_\Gamma\, \varepsilon(\beta_j) + \cdots + \sigma_{\ell+k}\alpha_{i,\ell+k} \,{}^{\mathsf{s}}\|_\Gamma\, \varepsilon(\beta_j)$. Then, the result follows from the distributivity of ${}^{\mathsf{s}}\|_\Gamma$ over $+$. In fact, the elements of $\alpha_{i,p} \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,p}$ belong to

$$\pi^{\sigma_p}(\alpha_0) \,{}^{\mathsf{s}}\|_\Gamma\, \pi^{\sigma_p}(\beta_0) = \pi^{\sigma_p}(\alpha_0 \,{}^{\mathsf{s}}\|_\Gamma\, \beta_0),$$

for $0 \le i \le n_1$, $0 \le j \le n_2$, and $1 \le p \le \ell$. Furthermore, $\alpha_{i,\ell+p} \,{}^{\mathsf{s}}\|_\Gamma\, \beta_j \in \pi^{\sigma_{\ell+p}}(\alpha_0) \,{}^{\mathsf{s}}\|_\Gamma\, (\pi(\beta_0) \cup \{\beta_0\}) \subseteq \pi^{\sigma_{\ell+p}}(\alpha_0 \,{}^{\mathsf{s}}\|_\Gamma\, \beta_0)$ and similarly for $\alpha_i \,{}^{\mathsf{s}}\|_\Gamma\, \beta_{j,\ell+p}$.

Now, let $\circ = {}^{\mathsf{a}}\|_\Gamma$. Then:

$$\begin{aligned}
\alpha_i \,{}^{\mathsf{a}}\|_\Gamma\, \beta_j \doteq\; & (\sigma_1\alpha_{i,1} + \cdots + \sigma_\ell\alpha_{i,\ell} + \sigma_{\ell+1}\alpha_{i,\ell+1} + \cdots + \sigma_{\ell+k}\alpha_{i,\ell+k} + \varepsilon(\alpha_i)) \\
& {}^{\mathsf{a}}\|_\Gamma\, (\sigma_1\beta_{j,1} + \cdots + \sigma_\ell\beta_{j,\ell} + \sigma_{\ell+1}\beta_{j,\ell+1} + \cdots + \sigma_{\ell+k}\beta_{j,\ell+k} + \varepsilon(\beta_j)) \\
& \doteq \sigma_1(\alpha_{i,1} \,{}^{\mathsf{a}}\|_\Gamma\, \beta_{j,1}) + \cdots + \sigma_\ell(\alpha_{i,\ell} \,{}^{\mathsf{a}}\|_\Gamma\, \beta_{j,\ell}) \\
& + \sigma_1(\alpha_{i,1} \,{}^{\mathsf{a}}\|_\Gamma\, \beta_j) + \cdots + \sigma_{\ell+k}(\alpha_{i,\ell+k} \,{}^{\mathsf{a}}\|_\Gamma\, \beta_j) \\
& + \sigma_1(\alpha_i \,{}^{\mathsf{a}}\|_\Gamma\, \beta_{j,1}) + \cdots + \sigma_{\ell+k}(\alpha_i \,{}^{\mathsf{a}}\|_\Gamma\, \beta_{j,\ell+k}) \\
& + \varepsilon(\alpha_i) \,{}^{\mathsf{a}}\|_\Gamma\, \sigma_1\beta_{j,1} + \cdots + \varepsilon(\alpha_i) \,{}^{\mathsf{a}}\|_\Gamma\, \sigma_{\ell+k}\beta_{j,\ell+k} \\
& + \sigma_1\alpha_{i,1} \,{}^{\mathsf{a}}\|_\Gamma\, \varepsilon(\beta_j) + \cdots + \sigma_{\ell+k}\alpha_{i,\ell+k} \,{}^{\mathsf{a}}\|_\Gamma\, \varepsilon(\beta_j) + \varepsilon(\alpha_i) \,{}^{\mathsf{a}}\|_\Gamma\, \varepsilon(\beta_j) \\
\doteq\; & \sigma_1(\alpha_{i,1} \,{}^{\mathsf{a}}\|_\Gamma\, \beta_{j,1} + \alpha_{i,1} \,{}^{\mathsf{a}}\|_\Gamma\, \beta_j + \alpha_i \,{}^{\mathsf{a}}\|_\Gamma\, \beta_{j,1}) + \cdots + \\
& \sigma_\ell(\alpha_{i,\ell} \,{}^{\mathsf{a}}\|_\Gamma\, \beta_{j,\ell} + \alpha_{i,\ell} \,{}^{\mathsf{a}}\|_\Gamma\, \beta_j + \alpha_i \,{}^{\mathsf{a}}\|_\Gamma\, \beta_{j,\ell}) + \sigma_{\ell+1}(\alpha_{i,\ell+1} \,{}^{\mathsf{a}}\|_\Gamma\, \beta_j + \alpha_i \,{}^{\mathsf{a}}\|_\Gamma\, \beta_{j,\ell+1}) \\
& + \cdots + \sigma_{\ell+k}(\alpha_{i,\ell+k} \,{}^{\mathsf{a}}\|_\Gamma\, \beta_j + \alpha_i \,{}^{\mathsf{a}}\|_\Gamma\, \beta_{j,\ell+k}) + \varepsilon(\alpha_i \,{}^{\mathsf{a}}\|_\Gamma\, \beta_j).
\end{aligned}$$

Again, the result follows from a similar argument as in the case of ${}^{\mathsf{s}}\|_\Gamma$. Also, the elements of $\alpha_{i,p} \,{}^{\mathsf{a}}\!\|_\Gamma\, \beta_{j,p}$ belong to

$$\begin{aligned}\pi^{\sigma_p}(\alpha_0 \,{}^{\mathsf{a}}\!\|_\Gamma\, \beta_0) = \pi^{\sigma_p}(\alpha_0) \,{}^{\mathsf{a}}\!\|_\Gamma\, \pi^{\sigma_p}(\beta_0) \\ \cup\, \pi^{\sigma_p}(\alpha_0) \,{}^{\mathsf{a}}\!\|_\Gamma\, (\pi(\beta_0) \cup \{\beta_0\}) \cup (\pi(\alpha_0) \cup \{\alpha_0\}) \,{}^{\mathsf{a}}\!\|_\Gamma\, \pi^{\sigma_p}(\beta_0),\end{aligned}$$

for $0 \leq i \leq n_1$, $0 \leq j \leq n_2$, and $1 \leq p \leq \ell$. Furthermore,

$$\alpha_{i,\ell+p} \,{}^{\mathsf{s}}\!\|_\Gamma\, \beta_j \in \pi^{\sigma_{\ell+p}}(\alpha_0) \,{}^{\mathsf{s}}\!\|_\Gamma\, (\pi(\beta_0) \cup \{\beta_0\}) \subseteq \pi^{\sigma_{\ell+p}}(\alpha_0 \,{}^{\mathsf{s}}\!\|_\Gamma\, \beta_0)$$

and similarly for $\alpha_i \,{}^{\mathsf{s}}\!\|_\Gamma\, \beta_{j,\ell+p}$. □

Example 4. *Let* $\Sigma = \{a, b\}$ *and consider* $\alpha = (b + ab + aab + abab) \,{}^{\mathsf{s}}\!\|_{\{a,b\}}\, (ab)^\star$. *We have*

$$\begin{aligned}\pi^a(\alpha) &= \{b \,{}^{\mathsf{s}}\!\|_{\{a,b\}}\, b(ab)^\star, ab \,{}^{\mathsf{s}}\!\|_{\{a,b\}}\, b(ab)^\star, bab \,{}^{\mathsf{s}}\!\|_{\{a,b\}}\, b(ab)^\star\} \\ \pi^b(\alpha) &= \{\varepsilon \,{}^{\mathsf{s}}\!\|_{\{a,b\}}\, (ab)^\star, ab \,{}^{\mathsf{s}}\!\|_{\{a,b\}}\, (ab)^\star\}, \\ \pi(\alpha) &= \pi^a(\alpha) \cup \pi^b(\alpha),\end{aligned}$$

and $|\pi(\alpha)| = 5$. *This is an improvement w.r.t. the definition of* π *for expressions with the intersection operator in* [2], *Example 12, for which* $|\pi(\alpha)| = 8$.

4. Partial Derivatives and Partial Derivative Automata

The notions of partial derivatives and partial derivative automata of standard regular expressions were introduced by Antimirov [1]. Champarnaud and Ziadi [9] showed that the partial derivative automaton and Mirkin's contruction are identical. Sulzmann and Thiemann [15] extended partial derivatives and the partial derivative automaton to regular expressions with synchronised shuffles. In this section, we recall those notions and relate the set of partial derivatives with the support defined in the previous section. The set of partial derivatives of an expression $\alpha \in \mathrm{RE}(\|)$ by a symbol $\sigma \in \Sigma$, denoted by $\partial_\sigma(\alpha)$, is defined inductively as follows.

$$\partial_\sigma(\emptyset) = \partial_\sigma(\varepsilon) = \emptyset, \quad \partial_\sigma(\alpha^\star) = \partial_\sigma(\alpha)\alpha^\star, \quad \partial_\sigma(\sigma') = \begin{cases}\{\varepsilon\} \text{ if } \sigma = \sigma', \\ \emptyset \text{ otherwise,}\end{cases}$$

$$\partial_\sigma(\alpha + \beta) = \partial_\sigma(\alpha) \cup \partial_\sigma(\beta), \quad \partial_\sigma(\alpha\beta) = \partial_\sigma(\alpha)\beta \cup \varepsilon(\alpha)\partial_\sigma(\beta),$$

$$\partial_\sigma(\alpha \,{}^{\mathsf{s}}\!\|_\Gamma\, \beta) = \begin{cases}\partial_\sigma(\alpha) \,{}^{\mathsf{s}}\!\|_\Gamma\, \partial_\sigma(\beta) \text{ if } \sigma \in \Gamma, \\ \partial_\sigma(\alpha) \,{}^{\mathsf{s}}\!\|_\Gamma\, \{\beta\} \cup \{\alpha\} \,{}^{\mathsf{s}}\!\|_\Gamma\, \partial_\sigma(\beta), \text{ otherwise,}\end{cases} \tag{9}$$

$$\partial_\sigma(\alpha \,{}^{\mathsf{a}}\!\|_\Gamma\, \beta) = \begin{cases}\partial_\sigma(\alpha) \,{}^{\mathsf{a}}\!\|_\Gamma\, \partial_\sigma(\beta) \cup \partial_\sigma(\alpha) \,{}^{\mathsf{a}}\!\|_\Gamma\, \{\beta\} \cup \{\alpha\} \,{}^{\mathsf{a}}\!\|_\Gamma\, \partial_\sigma(\beta) \text{ if } \sigma \in \Gamma, \\ \partial_\sigma(\alpha) \,{}^{\mathsf{a}}\!\|_\Gamma\, \{\beta\} \cup \{\alpha\} \,{}^{\mathsf{a}}\!\|_\Gamma\, \partial_\sigma(\beta), \text{ otherwise.}\end{cases}$$

As usual, the set of partial derivatives of $\alpha \in \mathrm{RE}(\|)$ w.r.t. a word $w \in \Sigma^\star$ is inductively defined by $\partial_\varepsilon(\alpha) = \{\alpha\}$ and $\partial_{w\sigma}(\alpha) = \partial_\sigma(\partial_w(\alpha))$, where, given a set $S \subseteq \mathrm{RE}(\|)$, $\partial_\sigma(S) = \bigcup_{\alpha \in S} \partial_\sigma(\alpha)$. Moreover, $\mathcal{L}(\partial_w(\alpha)) = \{\, w_1 \mid ww_1 \in \mathcal{L}(\alpha)\,\}$. Let $\partial(\alpha) = \bigcup_{w \in \Sigma^\star} \partial_w(\alpha)$, and $\partial^+(\alpha) = \bigcup_{w \in \Sigma^+} \partial_w(\alpha)$. The partial derivative automaton of $\alpha \in \mathrm{RE}(\|)$ is $\mathcal{A}_{\mathrm{PD}}(\alpha) = \langle \partial(\alpha), \Sigma, \delta_{\mathrm{PD}}, \{\alpha\}, F_{\mathrm{PD}}\rangle$, with $F_{\mathrm{PD}} = \{\, \beta \in \partial(\alpha) \mid \varepsilon(\beta) = \varepsilon\,\}$ and $\delta_{\mathrm{PD}}(\beta, \sigma) = \partial_\sigma(\beta)$, for $\beta \in \partial(\alpha)$, $\sigma \in \Sigma$.

We now relate the set of partial derivatives with the support from Proposition 3. The following lemma is essential to obtain Proposition 7.

Lemma 5. *For $\sigma, \tau \in \Sigma$ and $\alpha \in \mathrm{RE}(\|)$, we have $\pi^\sigma(\pi^\tau(\alpha)) \subseteq \pi^\sigma(\alpha)$.*

Proof. We proceed by structural induction on α. If α is a letter or equal to ε, then $\pi^\sigma(\pi^\tau(\alpha)) = \emptyset$ and the result is true. For $\alpha + \beta$, we have

$$\pi^\sigma(\pi^\tau(\alpha+\beta)) = \pi^\sigma(\pi^\tau(\alpha)) \cup \pi^\sigma(\pi^\tau(\beta)) \subseteq \pi^\sigma(\alpha) \cup \pi^\sigma(\beta) = \pi^\sigma(\alpha+\beta).$$

For $\alpha\beta$, we have

$$\begin{aligned}\pi^\sigma(\pi^\tau(\alpha\beta)) = \pi^\sigma(\pi^\tau(\alpha)\beta \cup \pi^\tau(\beta)) &\subseteq \pi^\sigma(\pi^\tau(\alpha))\beta \cup \pi^\sigma(\beta) \cup \pi^\sigma(\pi^\tau(\beta)) \\ &\subseteq \pi^\sigma(\alpha)\beta \cup \pi^\sigma(\beta) = \pi^\sigma(\alpha\beta).\end{aligned}$$

For $\alpha^\star$, we have

$$\begin{aligned}\pi^\sigma(\pi^\tau(\alpha^\star)) = \pi^\sigma(\pi^\tau(\alpha)\alpha^\star) &= \pi^\sigma(\pi^\tau(\alpha))\alpha^\star \cup \pi^\sigma(\alpha^\star) \\ &\subseteq \pi^\sigma(\alpha)\alpha^\star \cup \pi^\sigma(\alpha^\star) = \pi^\sigma(\alpha^\star).\end{aligned}$$

For $\alpha \,{}^{\mathsf{s}}\|_\Gamma\, \beta$ and $\tau \in \Gamma$, we have

$$\pi^\sigma(\pi^\tau(\alpha \,{}^{\mathsf{s}}\|_\Gamma\, \beta)) = \pi^\sigma(\pi^\tau(\alpha) \,{}^{\mathsf{s}}\|_\Gamma\, \pi^\tau(\beta)).$$

If $\sigma \in \Gamma$, then this expression evaluates to

$$\pi^\sigma(\pi^\tau(\alpha)) \,{}^{\mathsf{s}}\|_\Gamma\, \pi^\sigma(\pi^\tau(\beta)) \subseteq \pi^\sigma(\alpha) \,{}^{\mathsf{s}}\|_\Gamma\, \pi^\sigma(\beta) = \pi^\sigma(\alpha \,{}^{\mathsf{s}}\|_\Gamma\, \beta).$$

Otherwise, it evaluates to

$$\begin{aligned}&\pi^\sigma(\pi^\tau(\alpha)) \,{}^{\mathsf{s}}\|_\Gamma\, \pi(\pi^\tau(\beta)) \cup \ \pi^\sigma(\pi^\tau(\alpha)) \,{}^{\mathsf{s}}\|_\Gamma\, \pi^\tau(\beta) \\ &\qquad \cup \ \pi(\pi^\tau(\alpha)) \,{}^{\mathsf{s}}\|_\Gamma\, \pi^\sigma(\pi^\tau(\beta)) \cup \pi^\tau(\alpha) \,{}^{\mathsf{s}}\|_\Gamma\, \pi^\sigma(\pi^\tau(\beta)).\end{aligned}$$

By induction,

$$\pi(\pi^\tau(\beta)) = \bigcup_{\sigma\in\Sigma} \pi^\sigma(\pi^\tau(\beta)) \ \subseteq \bigcup_{\sigma\in\Sigma} \pi^\sigma(\beta) = \pi(\beta),$$

and in the same way $\pi(\pi^\tau(\alpha)) \subseteq \pi(\alpha)$. Thus, the result is a subset of

$$\begin{aligned}\pi^\sigma(\alpha) \,{}^{\mathsf{s}}\|_\Gamma\, \pi(\beta) \ \cup \ \pi^\sigma(\alpha) \,{}^{\mathsf{s}}\|_\Gamma\, \pi^\tau(\beta) \ \cup \ \pi(\alpha) \,{}^{\mathsf{s}}\|_\Gamma\, \pi^\sigma(\beta) \ \cup \ \pi^\tau(\alpha) \,{}^{\mathsf{s}}\|_\Gamma\, \pi^\sigma(\beta) \\ \subseteq \pi^\sigma(\alpha) \,{}^{\mathsf{s}}\|_\Gamma\, \pi(\beta) \ \cup \ \pi(\alpha) \,{}^{\mathsf{s}}\|_\Gamma\, \pi^\sigma(\beta) \\ \subseteq \pi^\sigma(\alpha \,{}^{\mathsf{s}}\|_\Gamma\, \beta).\end{aligned}$$

The case $\tau \notin \Gamma$ is proven in the same way. Finally, for $\alpha \,{}^{\mathsf{a}}\|_\Gamma\, \beta$ we only consider the case $\sigma, \tau \in \Gamma$, since the remaining ones are less involved. Note, that by induction,

$$\pi(\pi(\alpha)) = \bigcup_{\sigma\in\Sigma} \pi^\sigma\left(\bigcup_{\tau\in\Sigma} \pi^\tau(\alpha)\right) \subseteq \bigcup_{\sigma\in\Sigma}\left(\bigcup_{\tau\in\Sigma} \pi^\sigma(\alpha)\right) = \bigcup_{\sigma\in\Sigma} \pi^\sigma(\alpha) = \pi(\alpha),$$

while $\pi^\sigma(\pi(\alpha)) \subseteq \pi^\sigma(\alpha)$. Then,

$$\begin{aligned}\pi^\sigma(\pi^\tau(\alpha\,{}^{\mathsf{s}}\|_\Gamma\,\beta)) = {} & \pi^\sigma(\pi^\tau(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\pi(\beta)) \cup \pi^\sigma(\pi^\tau(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\beta)\\ & \cup\ \pi^\sigma(\pi(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\pi^\tau(\beta)) \cup \pi^\sigma(\alpha\,{}^{\mathsf{a}}\|_\Gamma\,\pi^\tau(\beta))\\ & \cup\ \pi^\sigma(\pi^\tau(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\pi^\tau(\beta)).\end{aligned}$$

It remains to show that all components of this union are subsets of

$$\begin{aligned}\pi^\sigma(\alpha\,{}^{\mathsf{s}}\|_\Gamma\,\beta) = {} & \pi^\sigma(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\pi(\beta) \cup \pi^\sigma(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\beta\\ & \cup\ \pi(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\pi^\sigma(\beta) \cup \alpha\,{}^{\mathsf{a}}\|_\Gamma\,\pi^\sigma(\beta)\\ & \cup\ \pi^\sigma(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\pi^\sigma(\beta).\end{aligned}$$

In fact,

$$\begin{aligned}\pi^\sigma(\pi^\tau(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\pi(\beta)) = {} & \pi^\sigma(\pi^\tau(\alpha))\,{}^{\mathsf{a}}\|_\Gamma\,\pi(\pi(\beta))\ \cup\ \pi^\sigma(\pi^\tau(\alpha))\,{}^{\mathsf{a}}\|_\Gamma\,\pi(\beta)\\ & \cup \pi(\pi^\tau(\alpha))\,{}^{\mathsf{a}}\|_\Gamma\,\pi^\sigma(\pi(\beta))\ \cup\ \pi^\tau(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\pi^\sigma(\pi(\beta))\\ & \cup\ \pi^\sigma(\pi^\tau(\alpha))\,{}^{\mathsf{a}}\|_\Gamma\,\pi^\sigma(\pi(\beta))\\ \subseteq {} & \ \pi^\sigma(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\pi(\beta)\ \cup\ \pi(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\pi^\sigma(\beta)\ \cup\ \pi^\sigma(\alpha)\,{}^{\mathsf{a}}\|_\Gamma\,\pi^\sigma(\beta)\\ \subseteq {} & \pi^\sigma(\alpha\,{}^{\mathsf{s}}\|_\Gamma\,\beta).\end{aligned}$$

The establishment of this relation for the remaining components is similar. □

Lemma 6. *For $\sigma \in \Sigma$ and $\alpha \in \mathrm{RE}(\|)$, we have $\partial_\sigma(\alpha) \subseteq \pi^\sigma(\alpha)$.*

Proof. By structural induction on α. □

Proposition 7. *For $w\sigma \in \Sigma^+$ and $\alpha \in \mathrm{RE}(\|)$, we have $\partial_{w\sigma}(\alpha) \subseteq \pi^\sigma(\alpha)$.*

Proof. By induction on $|w|$. For $w = \varepsilon$, the result follows from Lemma 6. Now, consider a word of the form $w\sigma$, where $w = w'\tau$. Then, $\partial_{w\sigma}(\alpha) = \partial_\sigma(\partial_w(\alpha)) \subseteq \partial_\sigma(\pi^\tau(\alpha))$. By Lemma 6 and Lemma 5, $\partial_\sigma(\pi^\tau(\alpha)) \subseteq \pi^\sigma(\pi^\tau(\alpha)) \subseteq \pi^\sigma(\alpha)$. □

As an immediate consequence, we have the following proposition establishing that the support of an expression is a superset of the set of states of its $\mathcal{A}_{\mathrm{PD}}$.

Proposition 8. *Given $\alpha \in \mathrm{RE}(\|)$, $\partial^+(\alpha) \subseteq \pi(\alpha)$.*

Example 9. *For α from Example 4, we have*

$$\begin{aligned}\partial^+(\alpha) = \{ & bab\,{}^{\mathsf{s}}\|_{\{a,b\}}\,b(ab)^\star, ab\,{}^{\mathsf{s}}\|_{\{a,b\}}\,b(ab)^\star,\ b\,{}^{\mathsf{s}}\|_{\{a,b\}}\,b(ab)^\star,\\ & ab\,{}^{\mathsf{s}}\|_{\{a,b\}}\,(ab)^\star,\ \varepsilon\,{}^{\mathsf{s}}\|_{\{a,b\}}\,(ab)^\star\}.\end{aligned}$$

Thus, in this case $\partial^+(\alpha) = \pi^a(\alpha) \cup \pi^b(\alpha) = \pi(\alpha)$. However, one can have $\partial^+(\alpha) \subsetneq \pi(\alpha)$. For instance,

$$\partial^+(ab\,{}^{\mathsf{s}}\|_{\{a,b\}}\,b^\star a) = \{b\,{}^{\mathsf{s}}\|_{\{a,b\}}\,\varepsilon\} \subsetneq \{b\,{}^{\mathsf{s}}\|_{\{a,b\}}\,\varepsilon, \varepsilon\,{}^{\mathsf{s}}\|_{\{a,b\}}\,b^\star a\} = \pi(\alpha).$$

5. Average Case Complexity

In this section, we will extensively use the Bachman-Landau notation, namely

$$f(n_1, \ldots, n_k) \underset{n_1, \ldots, n_k \to \infty}{\sim} g(n_1, \ldots, n_k) \text{ as } \lim_{n_k \to \infty} \cdots \lim_{n_1 \to \infty} \frac{f(n_1, \ldots, n_k)}{g(n_1, \ldots, n_k)} = 1.$$

Given some measure over the objects of a combinatorial class, $\mathcal{A}$, for each $n \in \mathbb{N}$, let a_n be the sum of the values of this measure for all objects of size n. Here we consider $\mathcal{A} = \mathrm{RE}(\|)$, and the measure (cost function) is the number of partial derivatives for expressions of size n. Let $A(z) = \sum_n a_n z^n$ be the corresponding generating function. We will use the notation $[z^n]A(z)$ for a_n. Seeing this generating function $A(z)$ as a complex analytic function, if it has a unique dominant singularity ρ, the study of the behaviour of $A(z)$ around ρ gives us access to the asymptotic form of its coefficients. For more details see [11]. In this section, we will make ample use of the notions and the techniques expounded in [6, 12]. Of particular relevance is [6] , which we here reproduce:

Theorem 10. *Let $G(z)$ be a generating function with non-negative integral coefficients, and $C(z, w) \in \mathbb{Q}[z, w]$ be such that $C(z, G(z)) = 0$. Assume that $G(z)$ has a unique dominant singularity, ρ. Then, if $\lim_{z \to \rho} G(z) = a \in \mathbb{R}$,*

$$[z^n]G(z) \underset{n \to \infty}{\sim} \frac{-b}{\Gamma(-\alpha)} \rho^{-n} n^{-\alpha - 1},$$

where α is the smallest non-zero exponent of the Puiseux expansion of $G(\rho - \rho s)$ with respect to the variable s, and in the case $\alpha = \frac{1}{2}$ (which is true in all the cases considered in this paper), b is given by:

$$b = \sqrt{\frac{2\rho \frac{\partial C}{\partial z}}{\frac{\partial^2 C}{\partial w^2}}}\Bigg|_{\substack{z=\rho \\ w=a}}.$$

5.1. *Average Number of Partial Derivatives by one Symbol*

We start by estimating the asymptotic average number of partial derivatives by one symbol for $\alpha \in \mathrm{RE}(\|)$, which corresponds to the expected number of transitions from a state in $\mathcal{A}_{\mathrm{PD}}$. For standard regular expressions and for large alphabets that value is known to be the constant 6 [14, 4]. The cases of the strong and arbitrarily synchronised shuffles are analysed individually, but the results will be essentially identical. For the strong synchronisation, we also consider the extreme cases of intersection and shuffle which were not considered before in the literature.

5.1.1. *Strong Synchronisation.*

Let Σ be the alphabet, and let $\Gamma \subseteq \Sigma$ with $\ell = |\Gamma|$. We set $k = |\Sigma \setminus \Gamma|$, and $m = k + \ell = |\Sigma|$. We consider regular expressions with the strong synchronisation operator over the alphabet Σ generated by the grammar (1) without expressions with the operator ${}^{\mathsf{a}}\|_\Gamma$. Moreover we will consider all operators using the same Γ.

We denote this set of expressions by RE($^{\mathsf{s}}\|_\Gamma$). The generating function, $R_m(z)$, whose coefficient $[z^n]R_m(z)$ is the cumulative number of those regular expressions of size n, satisfies

$$R_m(z) = (m+1)z + 3zR_m(z)^2 + zR_m(z).$$

Regular expressions that have ε in their language, denoted by α_ε, are unambiguously generated by the following grammar

$$\alpha_\varepsilon \to \varepsilon \mid (\alpha_\varepsilon + \alpha) \mid (\alpha_{\overline{\varepsilon}} + \alpha_\varepsilon) \mid (\alpha_\varepsilon \alpha_\varepsilon) \mid (\alpha^\star) \mid (\alpha_\varepsilon \,^{\mathsf{s}}\|_\Gamma\, \alpha_\varepsilon),$$

where $\alpha_{\overline{\varepsilon}}$ represents regular expressions that do not have ε in their language. Using $R_{\overline{\varepsilon},m}(z) = R_m(z) - R_{\varepsilon,m}(z)$, the generating function for α_ε satisfies

$$R_{\varepsilon,m}(z) = z + 2zR(z)R_{\varepsilon,m}(z) + zR_{\varepsilon,m}(z)^2 + zR_m(z).$$

Given $\tau \in \Sigma \setminus \Gamma$ and $\gamma \in \Gamma$, we denote by $\mathsf{t}(\alpha)$ and by $\mathsf{g}(\alpha)$ the cost functions for an upper bound of the cardinality of $\partial_\tau(\alpha)$ and the cardinality of $\partial_\gamma(\alpha)$, respectively. Using (9) we have

$$\begin{aligned}
\mathsf{t}(\varepsilon) &= \mathsf{t}(\sigma) = 0, \quad \sigma \neq \tau & \mathsf{t}(\tau) &= 1,\\
\mathsf{t}(\alpha + \beta) &= \mathsf{t}(\alpha) + \mathsf{t}(\beta), & \mathsf{t}(\alpha_\varepsilon \beta) &= \mathsf{t}(\alpha_\varepsilon) + \mathsf{t}(\beta),\\
\mathsf{t}(\alpha_{\overline{\varepsilon}} \beta) &= \mathsf{t}(\alpha_{\overline{\varepsilon}}), & \mathsf{t}(\alpha^\star) &= \mathsf{t}(\alpha),\\
\mathsf{t}(\alpha \,^{\mathsf{s}}\|_\Gamma\, \beta) &= \mathsf{t}(\alpha) + \mathsf{t}(\beta).
\end{aligned}$$

Note that, for $\alpha \,^{\mathsf{s}}\|_\Gamma\, \beta$ the number of partial derivatives by τ equals the sum of the number of derivatives by τ of both α and β. Thus,

$$\mathsf{t}(\alpha \,^{\mathsf{s}}\|_\Gamma\, \beta) = \mathsf{t}(\alpha) + \mathsf{t}(\beta).$$

The definition of $\mathsf{g}(\alpha)$ is analogous, except that $\mathsf{g}(\sigma) = 0$ if $\sigma \neq \gamma$ and $\mathsf{g}(\gamma) = 1$. Also, since $\partial_\gamma(\alpha \,^{\mathsf{s}}\|_\Gamma\, \beta) = \partial_\gamma(\alpha) \,^{\mathsf{s}}\|_\Gamma\, \partial_\gamma(\beta)$, we have

$$\mathsf{g}(\alpha \,^{\mathsf{s}}\|_\Gamma\, \beta) = \mathsf{g}(\alpha)\,\mathsf{g}(\beta).$$

The corresponding generating functions $T_m(z) = \sum_\alpha \mathsf{t}(\alpha) z^{|\alpha|}$ and $G_m(z) = \sum_\alpha \mathsf{g}(\alpha) z^{|\alpha|}$ satisfy, respectively,

$$\begin{aligned}
T_m(z) &= z + 5zT_m(z)R_m(z) + zT_m(z)R_{\varepsilon,m}(z) + zT_m(z),\\
G_m(z) &= z + 3zG_m(z)R_m(z) + zG_m(z)R_{\varepsilon,m}(z) + zG_m(z) + zG_m(z)^2.
\end{aligned}$$

The generating function for an upper bound of the cardinality of the set of partial derivatives by one symbol $\bigcup_{\sigma \in \Sigma} \partial_\sigma(\alpha)$ is given by

$$D_{k,\ell}(z) = kT_m(z) + \ell G_m(z).$$

To compute the asymptotic behaviour of the coefficients of $D_{k,\ell}(z)$, we proceed by dealing first with $G_m(z)$, and then with $T_m(z)$. Eliminating the auxiliary variables from the above equations [10], one obtains an algebraic curve given by

$C_m(z, w) \in \mathbb{Q}[z, w]$ such that $C_m(z, G_m(z)) = 0$. The polynomial $C_m(z, w)$ has degree 8 in w, and degree 6 in z. Using the techniques expounded in [6, 12], one finds that the minimal polynomial of the relevant singularity, let us call it $r_m(z)$, is a factor of the resultant of $C_m(z, w)$ and $\frac{\partial C_m}{\partial w}(z, w)$, which has 5 distinct irreducible factors. One of them is z, which of course cannot be $r_m(z)$, while another is

$$16m^2z^4 + 192mz^4 + 96mz^3 + 256z^4 + 8mz^2 + 128z^3 - 16z^2 - 8z + 1,$$

which is always positive for $z \in \mathbb{R}^+$, since

$$256z^4 + 128z^3 - 16z^2 - 8z + 1 = x^4 + 2x^3 - x^2 - 2x + 1 = (x^2 + x - 1)^2,$$

with $x = 4z$. One is left with three polynomials, and to find which is $r_m(z)$, one may proceed as follows. Looking at the (real part of the) graph of the curve $C_m(z, w)$, one realises that this curve has only one branch on the first quadrant, and that our generating function, being an increasing function, has a singularity that corresponds to the turning point with the lowest ordinate. In this way we found out that

$$r_m(z) = (12m + 11)z^2 + 2z - 1,$$

and therefore the singularity is

$$\rho_m = \frac{1}{1 + 2\sqrt{3m + 3}}, \tag{10}$$

while the minimal polynomial of $a_m = \lim_{z \to \rho_m} G_m(z)$ (see again [6, 12]) has degree 8:

$$\begin{aligned} &3w^8 + 6w^7 + (11 - 10m)w^6 + (10 - 14m)w^5 + (12 - 24m + 3m^2)w^4 + \\ &\qquad + (10 - 14m)w^3 + (11 - 10m)w^2 + 6w + 3. \end{aligned} \tag{11}$$

With the help of a plotting program that can deal with functions given in an implicit form, one can identify the root a_m of this polynomial pertaining to ρ_m, and then one can use Puiseux expansions to obtain the expansion of the appropriate a_m, which is, as $m \to \infty$,

$$\sqrt{3}\, m^{-\frac{1}{2}} + \frac{1}{12}m^{-1} + \frac{307\sqrt{3}}{64}\, m^{-\frac{3}{2}} + o(m^{-\frac{3}{2}}).$$

Using the techniques described in [6, 12], one gets:

$$[z^n]G_m(z) \underset{n,\, m \to \infty}{\sim} \frac{\sqrt{6}}{2\sqrt{\pi\, m}} \rho_m^{-n} n^{-\frac{3}{2}}. \tag{12}$$

With respect to T_m, the value of the singularity is the the same as for G_m, i.e. ρ_m, whereas the minimal polynomial of a_m is:

$$3m^2w^4 - 6mw^3 - (10m + 1)w^2 - 8w - 5, \tag{13}$$

and its Puiseux expansion is

$$\sqrt{3}\, m^{-\frac{1}{2}} + \frac{3}{4}m^{-1} + \frac{19\sqrt{3}}{64}\, m^{-\frac{3}{2}} + o(m^{-\frac{3}{2}}),$$

as $m \to \infty$. From this, one gets:

$$[z^n]T_m(z) \underset{n,\, m \to \infty}{\sim} \frac{5\sqrt{6}}{2\sqrt{\pi\, m}} \rho_m^{-n} n^{-\frac{3}{2}}. \tag{14}$$

Therefore,

$$[z^n]D_{k,\ell}(z) \underset{n,\, k,\, \ell \to \infty}{\sim} \frac{(k+5\ell)\sqrt{6}}{2\sqrt{\pi(k+\ell)}} \rho_m^{-n} n^{-\frac{3}{2}}. \tag{15}$$

Using the formulas in Section 5.1 of [2], with $s = m+1$, $u = 1$, $b = 3$, one obtains an estimate for the number of expressions, for large values of n and m,

$$[z^n]R_m(z) \underset{n \to \infty}{\sim} \frac{\sqrt{2-2\rho_m}}{12\rho_m\sqrt{\pi}} \rho_m^{-n} n^{-\frac{3}{2}} \underset{m \to \infty}{\sim} \sqrt{\frac{m}{6\pi}} \rho_m^{-n} n^{-\frac{3}{2}}. \tag{16}$$

Thus,

Proposition 11. *The average of the upper bound (here considered) of the number of partial derivatives by one symbol of an $\alpha \in \mathrm{RE}({}^{\mathsf{s}}\|_\Gamma)$ of size n is*

$$\frac{[z^n]D_{k,\ell}(z)}{[z^n]R_m(z)} \underset{n,\, k,\, \ell \to \infty}{\sim} \frac{6\sqrt{3}(k+5\ell)\rho_m}{\sqrt{k+\ell}\sqrt{1-\rho_m}}. \tag{17}$$

In particular, when $\ell = 0$, and thus $m = k$, one has

$$\lim_{k \to \infty} \frac{[z^n]D_{k,0}(z)}{[z^n]R_k(z)} = 3, \tag{18}$$

whereas when $k = 0$, and thus $m = \ell$, one has

$$\lim_{\ell \to \infty} \frac{[z^n]D_{0,\ell}(z)}{[z^n]R_\ell(z)} = 15, \tag{19}$$

and when $k = \ell = \frac{m}{2}$

$$\lim_{m \to \infty} \frac{[z^n]D_{\frac{m}{2},\frac{m}{2}}(z)}{[z^n]R_m(z)} = 9. \tag{20}$$

We recall that if $\ell = 0$, ${}^{\mathsf{s}}\|_\emptyset$ coincides with the shuffle operator, $⧢$; and if $k = 0$ ${}^{\mathsf{s}}\|_\Sigma$ coincides with intersection. The given results nicely relate with the estimated value for standard regular expressions mentioned above.

5.1.2. *Arbitrary Synchronisation.*

Now consider the set of expressions with the operator ${}^{\mathsf{a}}\|_\Gamma$ with Γ as above, i.e., $\mathrm{RE}({}^{\mathsf{a}}\|_\Gamma)$ and k, ℓ, m as in the previous section. The generating functions $R_m(z)$, $R_{\varepsilon,m}(z)$, and $T_m(z)$ coincide with the ones for $\mathrm{RE}({}^{\mathsf{s}}\|_\Gamma)$. In the definition of $\mathsf{g}(\alpha)$ we have

$$\mathsf{g}(\alpha \,{}^{\mathsf{a}}\|_\Gamma\, \beta) = \mathsf{g}(\alpha)\,\mathsf{g}(\beta) + \mathsf{g}(\alpha) + \mathsf{g}(\beta),$$

and, thus,

$$G_m(z) = z + 4zG_m(z)R_m(z) + zG_m(z)R_{\varepsilon,m}(z) + zG_m(z) + zG_m(z)^2.$$

In this case, the minimal polynomial for the singularity, σ_m, is

$$\begin{cases} (64m^2+579m+925)z^3+(84m+161)z^2-4(m+11)z-6, & \text{when } m \leq 6,\\ (12m+11)z^2+2z-1, & \text{otherwise,} \end{cases} \tag{21}$$

and, for $m \leq 6$,

$$\sigma_m = \frac{1}{4}m^{-\frac{1}{2}} - \frac{3}{32}m^{-1} - \frac{77}{256}m^{-\frac{3}{2}} + o(m^{-\frac{3}{2}}), \text{ as } m \to \infty, \tag{22}$$

while $\sigma_m = \rho_m$ for $m \geq 7$. It turns out that $a_m = 1$, for $1 \leq m \leq 6$, while for $m \geq 7$ one has

$$a_m = \frac{\sqrt{3}}{2}m^{-\frac{1}{2}} + \frac{3}{16}m^{-1} + \frac{77\sqrt{3}}{256}m^{-\frac{3}{2}} + o(m^{-\frac{3}{2}}), \text{ as } m \to \infty. \tag{23}$$

Using this, one arrives at the same result as in (12). Thus, one will obtain the same asymptotic estimates for the average size of $\bigcup_{\sigma\in\Sigma}\partial_\sigma(\alpha)$ as for $\mathrm{RE}(\,{}^{\mathsf{s}}\|_\Gamma\,)$.

5.2. *Average Size of the Support*

In this section we estimate upper bounds for the asymptotic average size of $\pi(\alpha)$, and thus of the average state complexity of $\mathcal{A}_{\mathrm{PD}}$. We treat the operators ${}^{\mathsf{s}}\|_\Gamma$ and ${}^{\mathsf{a}}\|_\Gamma$ seperately.

5.2.1. *Strong Synchronisation*

In this section we consider $\alpha \in \mathrm{RE}(\,{}^{\mathsf{s}}\|_\Gamma\,)$ with Γ used as in Section 5.1. Let $\Gamma \subseteq \Sigma$ with $|\Gamma| = \ell$ and $|\Sigma \setminus \Gamma| = k$, and $m = k + \ell$.

Let $\mathsf{p}(\alpha)$ be the cost function for an upper bound of the size of

$$\pi(\alpha) = \bigcup_{\gamma\in\Gamma}\pi^\gamma(\alpha) \cup \bigcup_{\tau\in\Sigma\setminus\Gamma}\pi^\tau(\alpha).$$

For computing $\mathsf{p}(\alpha)$, let $\mathsf{s}(\alpha)$ be the cost function for an upper bound of the size of $\pi^\gamma(\alpha)$, where $\gamma \in \Gamma$. Using (6), we have

$$\begin{aligned} \mathsf{s}(\varepsilon) &= 0 = \mathsf{s}(\sigma), \quad \text{for } \sigma \neq \gamma, & \mathsf{s}(\alpha+\beta) &= \mathsf{s}(\alpha)+\mathsf{s}(\beta),\\ \mathsf{s}(\gamma) &= 1, & \mathsf{s}(\alpha\beta) &= \mathsf{s}(\alpha)+\mathsf{s}(\beta),\\ \mathsf{s}(\alpha^\star) &= \mathsf{s}(\alpha), & \mathsf{s}(\alpha\,{}^{\mathsf{s}}\|_\Gamma\,\beta) &= \mathsf{s}(\alpha)\,\mathsf{s}(\beta). \end{aligned}$$

In the same way, let $\mathsf{u}(\alpha)$ be the cost function for an upper bound of the size of $\pi^\tau(\alpha)$, where $\tau \in \Sigma \setminus \Gamma$. We have:

$$\begin{aligned} \mathsf{u}(\varepsilon) &= 0 = \mathsf{u}(\sigma), \quad \text{for } \sigma \neq \tau, & \mathsf{u}(\alpha+\beta) &= \mathsf{u}(\alpha)+\mathsf{u}(\beta),\\ \mathsf{u}(\tau) &= 1, & \mathsf{u}(\alpha\beta) &= \mathsf{u}(\alpha)+\mathsf{u}(\beta),\\ \mathsf{u}(\alpha^\star) &= \mathsf{u}(\alpha), & \mathsf{u}(\alpha\,{}^{\mathsf{s}}\|_\Gamma\,\beta) &= \mathsf{p}(\alpha)\,\mathsf{u}(\beta)+\ell\,\mathsf{u}(\alpha)\,\mathsf{s}(\beta)+\mathsf{u}(\alpha)+\mathsf{u}(\beta), \end{aligned}$$

where in $\mathsf{u}(\alpha\,{}^{\mathsf{s}}\|_\Gamma\,\beta)$ we avoid to count twice $\mathsf{u}(\alpha)\,\mathsf{u}(\beta)$. Then,

$$\mathsf{p}(\alpha) = \ell\,\mathsf{s}(\alpha) + k\,\mathsf{u}(\alpha).$$

The generating functions for $\mathsf{s}(\alpha)$, $\mathsf{u}(\alpha)$, and $\mathsf{p}(\alpha)$, respectively $S_m(z)$, $U_m(z)$ and $P_m(z)$, satisfy the following equalities:

$$S_m(z) = z + 4zS_m(z)R_m(z) + zS_m(z) + zS_m(z)^2, \tag{24}$$
$$U_m(z) = z + 6zU_m(z)R_m(z) + zU_m(z)P_{k,\ell}(z) + \ell zS_m(z)U_m(z) + zU_m(z), \tag{25}$$
$$P_{k,\ell}(z) = \ell S_m(z) + kU_m(z). \tag{26}$$

Using the same procedure as described above, one obtains a polynomial $C_{k,\ell}(z, w) \in \mathbb{Q}[z, w]$, such that $C_{k,\ell}(z, P_{k,\ell}(z)) = 0$. This polynomial $C_{k,\ell}$ has degree 8 in w, and degree 6 in z.

For $k = \ell = \frac{m}{2}$, using graphical and numerical methods, as well as Puiseux expansions, one obtains that the relevant singularity, η_m, which is a root of a polynomial of degree 8, has the following asymptotic behaviour:

$$\eta_m \underset{m\to\infty}{\sim} \sqrt{2}\,\beta\, m^{-\frac{1}{2}}, \tag{27}$$

where $\beta \simeq 0.180866$ is the biggest root of the polynomial $1100z^4 + 8z^3 - 68z^2 + 1$.

Using the same techniques as above, one sees that:

$$[z^n]P_{\frac{m}{2},\frac{m}{2}}(z) \underset{n,\, m\to\infty}{\sim} \frac{\gamma\sqrt{m}}{2\sqrt{\pi}}\eta_m^{-n} n^{-\frac{3}{2}}, \tag{28}$$

where $\gamma \simeq 6.73978$, and therefore

$$\frac{[z^n]P_{\frac{m}{2},\frac{m}{2}}(z)}{[z^n]R_m(z)} \underset{n,\, m\to\infty}{\sim} \sqrt{\frac{3}{2}}\,\gamma \left(\frac{\rho_m}{\eta_m}\right)^n, \tag{29}$$

$$\lim_{m\to\infty} \frac{\rho_m}{\eta_m} = \frac{1}{2\sqrt{6}\,\beta} \simeq 1.12859. \tag{30}$$

Proposition 12. *For large values of m and n, and $k = \ell$, an upper bound for the average number of states of $\mathcal{A}_{\mathrm{PD}}(\alpha)$ for $\alpha \in \mathrm{RE}(\,^{\mathsf{s}}\|_\Gamma\,)$ is $(1.12859 + o(1))^n$.*

When $k = 0$, i.e., in case of intersection, one obtains a simpler polynomial $C_\ell(z, w)$ for the corresponding generating function, namely:

$$3z^2w^4 + 2\ell z(z-1)w^3 + \ell^2((16\ell + 21)z^2 + 2z - 1)w^2 + 2\ell^3 z(z-1)w + 3\ell^4 z^2. \tag{31}$$

The singularity is the same ρ_ℓ as above, and

$$a_\ell = \frac{\ell\sqrt{2\ell - 1 - 2\sqrt{\ell^2 - \ell - 2}}}{\sqrt{3}}.$$

This yields

$$[z^n]P_{0,\ell}(z) \underset{n,\, \ell\to\infty}{\sim} \sqrt{\frac{3}{2}\frac{\ell}{\pi}}\rho_\ell^{-n} n^{-\frac{3}{2}}. \tag{32}$$

Proposition 13. *With the notations introduced above, the average size of the number of states in the partial derivative automata for a (standard) regular expression with intersection is asymptotically*

$$\frac{[z^n]P_{0,\ell}(z)}{[z^n]R_\ell(z)} \underset{n,\,\ell\to\infty}{\sim} 3. \tag{33}$$

This result is a surprising improvement of the previous upper bound of $(1.056 + o(1))^n$ given in [2] but is compatible with experimental values given in that paper and discussed in Section 6.

Finally, when $\ell = 0$, i.e., in the case of shuffle, a polynomial $C_k(z, w)$ for the corresponding generating function is:

$$z^2w^4 + ((14k + 11)z^2 + 2z - 1)w^2 + k^2z^2. \tag{34}$$

The singularity is now $\xi_k = \frac{1}{1+2\sqrt{4k+3}}$, and $a_k = \sqrt{k}$. This yields:

$$[z^n]P_{k,0}(z) \underset{n,\,k\to\infty}{\sim} \sqrt{\frac{2k}{\pi}}\xi_k^{-n}n^{-\frac{3}{2}}, \tag{35}$$

and therefore, we obtain the same result as in [5], but using different techniques.

Proposition 14. *With the notations introduced above, the average size of the number of states in the partial derivative automata for a (standard) regular expression with shuffle is asymptotically*

$$\frac{[z^n]P_{k,0}(z)}{[z^n]R_k(z)} \underset{n,\,k\to\infty}{\sim} 2\sqrt{3}\left(\frac{\rho_k}{\xi_k}\right)^n \underset{k\to\infty}{\sim} 2\sqrt{3}\left(\frac{4}{3}\right)^{\frac{n}{2}}. \tag{36}$$

5.2.2. *Arbitrary Synchronisation*

We now consider expressions with the operator ${}^{\mathsf{a}}\|_\Gamma$, RE(${}^{\mathsf{a}}\|_\Gamma$). As in the previous section, let $\mathsf{p}(\alpha)$ be the cost function for an upper bound of the size of $\pi(\alpha)$, let $\mathsf{s}(\alpha)$ be the upper bound of the size of $\pi^\gamma(\alpha)$ for $\gamma \in \Gamma$, and let $\mathsf{u}(\alpha)$ be the cost function of an upper bound of the size of $\pi^\tau(\alpha)$ where $\tau \in \Sigma \setminus \Gamma$. The function u is the same as for RE(${}^{\mathsf{s}}\|_\Gamma$). The definition of s is now

$$\mathsf{s}(\alpha\,{}^{\mathsf{a}}\|_\Gamma\,\beta) = \mathsf{p}(\alpha)\,\mathsf{s}(\beta) + k\,\mathsf{s}(\alpha)\,\mathsf{u}(\beta) + \mathsf{s}(\alpha) + \mathsf{s}(\beta) + \mathsf{s}(\alpha)\,\mathsf{s}(\beta)$$

and thus the generating functions for s is

$$S_m(z) = z + 6zS_m(z)R_m(z) + zS_m(z) + zS_m(z)^2 + zP_m(z)S_m(z) + zkS_m(z)U_m(z).$$

The definitions of the generating functions for u and p, $U_m(z)$ and $P_m(z)$, coincide with the ones for RE(${}^{\mathsf{s}}\|_\Gamma$), c.f. (26).

In the following we compute the asymptotic behaviour of $P_m(z)$. For $k = \ell = \frac{m}{2}$, using, as above, graphical and numerical methods, as well as Puiseux expansions, one obtains that the relevant singularity, η_m, which is a root of a polynomial of degree 8, has the following asymptotic behaviour:

$$\eta_m \underset{m\to\infty}{\sim} \frac{1}{3\sqrt{2}}\, m^{-\frac{1}{2}}. \tag{37}$$

This yields

$$[z^n]P_{\frac{m}{2},\frac{m}{2}}(z) \underset{n,\, m\to\infty}{\sim} \sqrt{\frac{m}{2\pi}}\,\eta_m^{-n} n^{-\frac{3}{2}}, \tag{38}$$

and therefore

$$\frac{[z^n]P_{\frac{m}{2},\frac{m}{2}}(z)}{[z^n]R_m(z)} \underset{n,\, m\to\infty}{\sim} \sqrt{3}\left(\frac{\rho_m}{\eta_m}\right)^n, \tag{39}$$

$$\lim_{m\to\infty}\frac{\rho_m}{\eta_m} = \sqrt{\frac{3}{2}} \simeq 1.22474. \tag{40}$$

Proposition 15. *For large values of m and n, and $k = \ell$, an upper bound for the average number of states of $\mathcal{A}_{\mathrm{PD}}(\alpha)$ for $\alpha \in \mathrm{RE}({}^{\mathsf{a}}\|_{\Gamma})$ is $(1.22474 + o(1))^n$.*

6. Experimental Results

In Table 1 we compare some experimental results obtained in [2] with analytic estimates given in this paper when considering regular expressions with ${}^{\mathsf{s}}\|_{\Sigma}$, i.e., intersection. To obtain those results regular expressions were uniformly random generated using a version of the grammar for $\mathrm{RE}(\cap)$ in prefix notation. For each size $n \in \{25, 50, 100, 150, 200, 300\}$ and alphabet size $\ell \in \{2, 10\}$ samples of 10000 regular expressions were generated. For each sample we computed the average value of several measures. The values $|\alpha|_{\Sigma}$ and $|\alpha|_{\cap}$ give, respectively, the number of alphabetic symbols and the number of operators $\cap$ in α. The column labelled with $|\partial(\alpha)|$, which is identical to $|\pi(\alpha)\cup\{\alpha\}|$, indicates the average number of states of the partial derivative automaton. The asymptotic values corresponding to the entries in this column are displayed in the last column, which contains the ratio of the coefficients of the formal series obtained in this paper. As one can verify, the values in the two columns are quite close, already for small expressions. Furthermore, we can conclude that on average the size of the $\mathcal{A}_{\mathrm{PD}}$ for regular expressions with intersection is small.

7. Conclusions

In this paper we studied the average number of partial derivatives of regular expressions with different synchronised shuffle operators. Partial derivatives have been extended to expressions containing strongly, weakly, and arbitrarily synchronised shuffle operators [3]. Here we estimated their number for expressions with the strong operator ${}^{\mathsf{s}}\|_{\Gamma}$ ($\Gamma \subseteq \Sigma$), as well as with the arbitrary operator ${}^{\mathsf{a}}\|_{\Gamma}$. The case of ${}^{\mathsf{s}}\|_{\Sigma}$ corresponds to intersection $\cap$, and ${}^{\mathsf{s}}\|_{\emptyset}$ to interleaving $\sqcup\!\sqcup$. In Proposition 11 we

Table 1. Experimental results and analytic estimations for expressions with $\cap$.

$m = \ell$	$n = \lvert\alpha\rvert$	$\lvert\alpha\rvert_\Sigma$	$\lvert\alpha\rvert_\cap$	$\lvert\partial(\alpha)\rvert$	$\frac{[z^n]P_{0,\ell}(z)}{[z^n]R_\ell(z)}$
2	25	7.46	3.41	2.47	2.295
	50	14.60	6.96	2.77	3.101
	100	28.88	14.14	3.07	4.138
	150	43.12	21.19	3.16	4.854
	200	57.44	28.41	3.22	5.417
	300	86.00	42.73	3.28	6.295
5	25	9.73	3.56	2.42	2.333
	50	19.04	7.28	2.52	2.812
	100	37.66	14.72	2.65	3.238
	150	56.34	22.16	2.71	3.448
	200	74.96	29.64	2.68	3.575
	300	112.19	44.61	2.72	3.725
10	25	10.90	3.63	2.38	2.332
	50	21.33	7.46	2.47	2.698
	100	42.26	15.16	2.50	2.897
	150	63.14	22.87	2.58	3.113
	200	84.13	30.43	2.58	3.185
	300	125	45.68	2.58	3.264

presented an upper bound for the average number of partial derivatives by just one symbol of the alphabet for expressions with $^{\mathsf{s}}\|_\Gamma$. It turned out that the same bounds hold for expressions with $^{\mathsf{a}}\|_\Gamma$. For the number of all partial derivatives we needed to consider new rules for computing the support π of an expression. Proposition 12 gives an upper bound for the average size of $\pi(\alpha)$ for the operator $^{\mathsf{s}}\|_\Gamma$. In particular, we show, in Proposition 13, that for intersection this upper bound is 3. This is a huge improvement of the known upper bound, but is corroborated by experimental results as shown in Section 6. Finally, in Proposition 15 we present an upper bound for the size of the support for the operator $^{\mathsf{a}}\|_\Gamma$.

The definition of partial derivatives for expressions with the weakly synchronised shuffle relies on an additional operator parameterised by two sets of alphabet symbols [15, 8]. As a consequence, the asymptotic analysis in this case is more complicated. However, it follows from the definitions presented in the papers above that $|\partial(\alpha[\mathsf{s}/\mathsf{w}]| \leq |\partial(\alpha)| \leq |\partial(\alpha[\mathsf{a}/\mathsf{w}]|$, whenever α is an expression with the weak synchronised operator $^{\mathsf{w}}\|_\Gamma$, and $\alpha[\mathsf{s}/\mathsf{w}]$ and $\alpha[\mathsf{a}/\mathsf{w}]$ are obtained by replacing in α each occurrence of $^{\mathsf{w}}\|_\Gamma$ by $^{\mathsf{s}}\|_\Gamma$, and by $^{\mathsf{a}}\|_\Gamma$, respectively. Therefore, the average complexity for the weak operator lies between those of the strong and of the arbitrary operator. In particular, the estimate of an upper bound for the average size of $\partial(\alpha)$ in the case of $^{\mathsf{a}}\|_\Gamma$ is also an upper bound for $^{\mathsf{w}}\|_\Gamma$.

References

[1] Antimirov, V.M.: Partial derivatives of regular expressions and finite automaton constructions. Theoret. Comput. Sci. **155**(2), 291–319 (1996)

[2] Bastos, R., Broda, S., Machiavelo, A., Moreira, N., Reis, R.: On the average complexity of partial derivative automata for semi-extended expressions. J. Autom. Lang. Comb. **22**(1–3), 5–28 (2017)

[3] ter Beek, M.H., Martín-Vide, C., Mitrana, V.: Synchronized shuffles. Theoret. Comput. Sci. (341), 263–275 (2005)

[4] Broda, S., Machiavelo, A., Moreira, N., Reis, R.: On the average size of Glushkov and partial derivative automata. Int. J. Found. Comput. S. **23**(5), 969–984 (2012)

[5] Broda, S., Machiavelo, A., Moreira, N., Reis, R.: Automata for regular expressions with shuffle. Inf. Comput. **259**(2), 162–173 (2018)

[6] Broda, S., Machiavelo, A., Moreira, N., Reis, R.: Analytic combinatorics and descriptional complexity of regular languages on average. ACM SIGACT News **51**(1), 38–56 (March 2020)

[7] Broda, S., Machiavelo, A., Moreira, N., Reis, R.: Average complexity of partial derivatives for synchronised shuffle expressions. In: Nagy, B. (ed.) CIAA 2023. LNCS, vol. 14151, pp. 103–115. Springer (2023), `https://doi.org/10.1007/978-3-031-40247-0\_7`

[8] Broda, S., Machiavelo, A., Moreira, N., Reis, R.: Location automata for synchronised shuffle expressions. J. Log. Algebr. Methods Program. **133** (2023)

[9] Champarnaud, J.M., Ziadi, D.: From Mirkin's prebases to Antimirov's word partial derivatives. Fundam. Inform. **45**(3), 195–205 (2001)

[10] Cox, D.A., John Little, D.O.: Ideals, Varieties, and Algorithms. Springer, 4th edn. (2018)

[11] Flajolet, P., Sedgewick, R.: Analytic Combinatorics. CUP (2008)

[12] Konstantinidis, S., Machiavelo, A., Moreira, N., Reis, R.: On the average state complexity of partial derivative transducers. In: Chatzigeorgiou, A., et al. (eds.) SOFSEM 2020. LNCS, vol. 12011, pp. 174–186. Springer (2020)

[13] Mirkin, B.G.: An algorithm for constructing a base in a language of regular expressions. Engineering Cybernetics **5**, 51—57 (1966)

[14] Nicaud, C.: Average state complexity of operations on unary automata. In: Kurylowski, M., et al. (eds.) 24th MFCS. LNCS, vol. 1672, pp. 231–240. Springer (1999)

[15] Sulzmann, M., Thiemann, P.: Derivatives and partial derivatives for regular shuffle expressions. J. Comput. System Sci. **104**, 323–341 (2019)