On the State Complexity of Partial Derivative Automata for Regular Expressions with Intersection*

Rafaela Bastos, Sabine Broda, António Machiavelo,
Nelma Moreira, and Rogério Reis

CMUP & Faculdade de Ciências da Universidade do Porto, Portugal
{rrbastos,sbb}@dcc.fc.up.pt,ajmachia@fc.up.pt,
{nam,rvr}@dcc.fc.up.pt

Abstract. Extended regular expressions (with complement and intersection) are used in many applications due to their succinctness. In particular, regular expressions extended with intersection only (also called semi-extended) can already be exponentially smaller than standard regular expressions or equivalent nondeterministic finite automata (NFA). For practical purposes it is important to study the average behaviour of conversions between these models. In this paper, we focus on the conversion of regular expressions with intersection to nondeterministic finite automata, using partial derivatives and the notion of support. First, we give a tight upper bound of $2^{O(n)}$ for the worst-case number of states of the resulting partial derivative automaton, where n is the size of the expression. Using the framework of analytic combinatorics, we then establish an upper bound of $(1.056+o(1))^n$ for its asymptotic average-state complexity, which is significantly smaller than the one for the worst case.

1 Introduction

Regular expressions with additional operators are used in applications such as programming languages [12], XML processing [23], or runtime verification [22]. Most of these operators do not increase their language expressive power but lead to gains in the succinctness of the representation. This is the case for intersection. For regular expressions with intersection ($\mathsf{RE}_\cap$) (or semi-extended), several computational complexity decision problems, such as membership, equivalence and emptiness, were studied by various authors. Petersen [21] has shown that the membership problem is LOGCFL-complete, while for standard regular expressions (RE) it is NL-complete [19]. Fürer [14] has proved that inequivalence and non-empty complement are EXPSPACE-complete, which contrasts with the PSPACE-completeness of these problems for RE. The complexity of the conversions from regular expressions with intersection to standard regular expressions,

* 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.

and to finite automata, were recently studied by Gelade and Neven [16], Gruber and Holzer [18], and Gelade [15]. The conversion from $\mathsf{RE}_\cap$ to RE or to nondeterministic finite automata (NFA) is exponential and it is double exponential to deterministic finite automata (DFA). The conversion from $\alpha \in \mathsf{RE}_\cap$ to a DFA can be accomplished using Brzozowski's derivatives [8]. From RE to NFA a standard algorithm is the partial derivative automaton construction ($\mathcal{A}_{pd}$) introduced by Antimirov [1], which coincides with the resolution of systems of equations by Mirkin [20]. The average complexity of these conversions was recently studied using the framework of analytic combinatorics [4,5], and also their extension to regular expressions with shuffle [7]. For these studies, Mirkin's construction is essential as it provides inductive definitions that can be used to obtain generating functions.

Caron et al. [9] extended the $\mathcal{A}_{pd}$ to regular expressions with both intersection and complement (extended regular expressions)[1]. In their approach a partial derivative is a set of sets of expressions (akin a disjunctive normal form), whereas here it is simply a set of expressions. In the worst-case, their approach also leads to NFAs that can be exponentially larger than the original expressions. Moreover, considering sets of sets of expressions would turn the analytic combinatoric analysis much harder.

In this paper we show that for $\mathsf{RE}_\cap$, Mirkin's construction can lead to automata not initially connected and thus larger than the ones built by Antimirov's construction. However, the two constructions can produce identical NFAs. We present an exponential worst-case upper bound which is tight for both. Using the framework of analytic combinatorics, we give an upper bound for the asymptotic average-state complexity for the Mirkin's construction, which turns out to be much smaller than the worst-case bound. This also means that Antimirov's construction is asymptotically and on average much smaller than the worst-case upper bound.

2 Regular Expressions with Intersection

Let $\Sigma = \{a_1, \ldots, a_k\}$ be an *alphabet* of size k. A *word* over Σ is a finite sequence of symbols of Σ. The *empty word* is denoted by ε. The set $\Sigma^\star$ is the set of all words over Σ. A *language* over Σ is a subset of $\Sigma^\star$. The set $\mathsf{RE}_\cap$ of *regular expressions with intersection* over Σ contains the expression $\emptyset$ and all terms generated by the following grammar:

$$\alpha \to \varepsilon \mid a \mid (\alpha + \alpha) \mid (\alpha \cdot \alpha) \mid (\alpha \cap \alpha) \mid (\alpha^\star) \qquad (a \in \Sigma), \tag{1}$$

where the operator $\cdot$ (concatenation) is often omitted. Parenthesis can also be omitted considering the following precedences for the operators: $\star > \cdot > \cap > +$. The size of a regular expression $\alpha \in \mathsf{RE}_\cap$ is denoted by $\|\alpha\|$ and defined as the number of occurrences of symbols (parenthesis not counted) in α. Similarly, $|\alpha|_\Sigma$ denotes the number of occurrences of alphabet symbols in α, and $|\alpha|_\cap$ the

[1] And a more general framework is also reported in [10].

number of occurrences of the binary operator $\cap$. The language $\mathcal{L}(\alpha)$ for $\alpha \in \mathsf{RE}_\cap$ is defined as usual, with $\mathcal{L}(\alpha \cap \beta) = \mathcal{L}(\alpha) \cap \mathcal{L}(\beta)$. We say that two regular expressions $\alpha, \beta \in \mathsf{RE}_\cap$ are *equivalent*, if $\mathcal{L}(\alpha) = \mathcal{L}(\beta)$, and write $\alpha \doteq \beta$ in this case. For a set $S \subseteq \mathsf{RE}_\cap$, the language of S is defined as $\mathcal{L}(S) = \bigcup_{\alpha \in S} \mathcal{L}(\alpha)$. The notion of equivalence extends naturally to sets of regular expressions. The *left-quotient* of a language $\mathcal{L}$ w.r.t. a word $w \in \Sigma^\star$ is defined as $w^{-1}\mathcal{L} = \{\ x \mid wx \in \mathcal{L}\ \}$. The algebraic structure $(\mathsf{RE}_\cap, +, \cdot, \emptyset, \varepsilon)$ constitutes an idempotent semiring, that with the unary operator $\star$ is a Kleene algebra. Antimirov and Mosses [2] presented a complete and sound axiomatization for $\mathsf{RE}_\cap$, where the binary operator $\cap$ is idempotent, commutative, associative, distributes over $+$, and also satisfies the following axioms, where $a_i, a_j \in \Sigma$:

$$
\begin{array}{ll}
(\varepsilon \cap \beta) \doteq \emptyset \wedge (\alpha \doteq \beta\alpha + \gamma) \Rightarrow \alpha \doteq \beta^\star\gamma, & \varepsilon \cap \alpha^\star \doteq \varepsilon, \\
\varepsilon \cap (\alpha\beta) \doteq (\varepsilon \cap \alpha) \cap \beta, & \varepsilon \cap a_i \doteq \emptyset \cap \alpha \doteq \emptyset, \\
(a_i\alpha) \cap (a_j\beta) \doteq (a_i \cap a_j)(\alpha \cap \beta), & a_i \cap a_j \doteq \emptyset \quad (a_i \neq a_j), \\
(\alpha a_i) \cap (\beta a_j) \doteq (\alpha \cap \beta)(a_i \cap a_j), & \alpha + (\alpha \cap \beta) \doteq \alpha.
\end{array}
$$

With the usual abuse of notation, define the function $\varepsilon : \mathsf{RE}_\cap \to \{\emptyset, \varepsilon\}$ by $\varepsilon(\alpha) = \varepsilon$ if $\varepsilon \in \mathcal{L}(\alpha)$, and $\varepsilon(\alpha) = \emptyset$ otherwise. The methods developed in Sections 3 and 4 are syntactical and aim at building automata equivalent to a given regular expression. To ensure the finiteness of the constructions it is not necessary to consider regular expressions modulo any of the above properties[2]. However, in some examples, for the sake of succinctness, we also consider regular expressions modulo the identities of $\cdot$ and $+$. Note that this does not affect the upper bounds of the number of states, both in the worst and in the average case.

3 Automata and Systems of Equations

We first recall the definition of a nondeterministic finite automaton (NFA) as 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, $\delta : S \times \Sigma \to 2^S$ the transition function, and $F \subseteq S$ a set of final states. The *language* of $\mathcal{A}$ is $\mathcal{L}(\mathcal{A}) = \{w \in \Sigma^\star \mid \delta(S_0, w) \cap F \neq \emptyset\}$. The *right language* of a state s, denoted by $\mathcal{L}_s$, is the language accepted by $\mathcal{A}$ if we take $S_0 = \{s\}$. It is well known that, for each n-state NFA $\mathcal{A}$, over $\Sigma = \{a_1, \ldots, a_k\}$, having right languages $\mathcal{L}_1, \ldots, \mathcal{L}_n$, it is possible to associate a system of linear language equations

$$\mathcal{L}_i = a_1\mathcal{L}_{1i} \cup \cdots \cup a_k\mathcal{L}_{ki} \cup \varepsilon(\mathcal{L}_i), \text{ for } i \in [1, n],$$

where $\mathcal{L}_{ji} = \bigcup_{l \in \delta(i, a_j)} \mathcal{L}_l$ and $\mathcal{L}(\mathcal{A}) = \bigcup_{i \in S_0} \mathcal{L}_i$. In the same way, it is possible to associate to each regular expression a system of equations. We here extend Mirkin's contruction to regular expressions with intersection.

Definition 1. *Consider $\alpha_0 \in \mathsf{RE}_\cap$ over $\Sigma = \{a_1, \ldots, a_k\}$. A* support *of α_0 is a set $\{\alpha_1, \ldots, \alpha_n\}$ of regular expressions with intersection that satisfies a system*

[2] As is the case, for instance, for Brzozowski DFA or Caron et al approach.

of equations

$$\alpha_i \doteq a_1\alpha_{1i} + \cdots + a_k\alpha_{ki} + \varepsilon(\alpha_i) \qquad i \in [0, n], \tag{2}$$

for some $\alpha_{1i}, \ldots, \alpha_{ki}$*, where each* $\alpha_{j,i}$ *is a (possibly empty) sum of elements in* $\{\alpha_1, \ldots, \alpha_n\}$.

It is clear that the existence of a support of α implies the existence of an NFA that accepts the language of α.

A support for a regular expression $\alpha \in \mathsf{RE}_\cap$ can be computed using the function $\pi : \mathsf{RE}_\cap \to 2^{\mathsf{RE}_\cap}$ defined below. First, we define some operations on sets of regular expressions. Given $S, T \subseteq \mathsf{RE}_\cap$ and $\beta \in \mathsf{RE}_\cap$, $S\beta = \{\ \alpha\beta \mid \alpha \in S\ \}$ and $S \cap T = \{\ \alpha \cap \beta \mid \alpha \in S, \beta \in T\ \}$. Note, in particular, that $\mathcal{L}(S \cap T) = \mathcal{L}(S) \cap \mathcal{L}(T)$.

Definition 2. *Given* $\alpha \in \mathsf{RE}_\cap$*, the set* $\pi(\alpha)$ *is inductively defined by:*

$$\begin{array}{ll} \pi(\emptyset) = \pi(\varepsilon) = \emptyset, & \pi(\alpha + \beta) = \pi(\alpha) \cup \pi(\beta), \\ \pi(a) = \{\varepsilon\} \quad (a \in \Sigma), & \pi(\alpha\beta) = \pi(\alpha)\beta \cup \pi(\beta), \\ \pi(\alpha^\star) = \pi(\alpha)\alpha^\star, & \pi(\alpha \cap \beta) = \pi(\alpha) \cap \pi(\beta). \end{array}$$

Proposition 3. *If* $\alpha \in \mathsf{RE}_\cap$*, then* $\pi(\alpha)$ *is a support of* α.

Proof. We will proceed by induction on the structure of α. The proof for all cases, excluding $\alpha \cap \beta$, can be found in [20,11,4]. Let $\pi(\alpha_0) = \{\alpha_1, \ldots, \alpha_n\}$ and $\pi(\beta_0) = \{\beta_1, \ldots, \beta_m\}$ be a support of α_0 and β_0, respectively. Thus,

$$\alpha_i \doteq a_1\alpha_{1i} + \cdots + a_k\alpha_{ki} + \varepsilon(\alpha_i), \quad \text{for } i = 0, \ldots, n$$

and

$$\beta_j \doteq a_1\beta_{1j} + \cdots + a_k\beta_{kj} + \varepsilon(\beta_j), \text{ for } j = 1, \ldots, m,$$

where, for all $l = 1, \ldots, k$, α_{li} and β_{lj} are linear combinations of elements of $\pi(\alpha_0)$ and $\pi(\beta_0)$, respectively. We want to prove that $\pi(\alpha_0 \cap \beta_0)$ is a support for $\alpha_0 \cap \beta_0$. For $i = 0, \ldots, n$ and $j = 0, \ldots, m$, and using the axioms for $\cap$, we have

$$\begin{aligned} \alpha_i \cap \beta_j \doteq & (a_1\alpha_{1i} + \cdots + a_k\alpha_{ki} + \varepsilon(\alpha_i)) \cap (a_1\beta_{1j} + \cdots + a_k\beta_{kj} + \varepsilon(\beta_j)) \\ \doteq & (a_1\alpha_{1i} \cap a_1\beta_{1j}) + \cdots + (a_1\alpha_{1i} \cap a_k\beta_{kj}) + (a_1\alpha_{1i} \cap \varepsilon(\beta_j)) + \\ & \ldots + (a_k\alpha_{ki} \cap a_1\beta_{1j}) + \cdots + (a_k\alpha_{ki} \cap a_k\beta_{kj}) + (a_k\alpha_{ki} \cap \varepsilon(\beta_j)) + \\ & \ldots + (\varepsilon(\alpha_i) \cap a_1\beta_{1j}) + \cdots + (\varepsilon(\alpha_i) \cap a_k\beta_{kj}) + (\varepsilon(\alpha_i) \cap \varepsilon(\beta_j)) \\ \doteq & (a_1 \cap a_1)(\alpha_{1i} \cap \beta_{1j}) + \cdots + (a_k \cap a_k)(\alpha_{ki} \cap \beta_{kj}) + (\varepsilon(\alpha_i) \cap \varepsilon(\beta_j)) \\ \doteq & a_1(\alpha_{1i} \cap \beta_{1j}) + \cdots + a_k(\alpha_{ki} \cap \beta_{kj}) + \varepsilon(\alpha_i \cap \beta_j). \end{aligned}$$

For each $l = 1, \ldots, k$, we know that $\alpha_{li} = \sum_{i' \in I_{li}} \alpha_{i'}$ and $\beta_{lj} = \sum_{j' \in J_{lj}} \beta_{j'}$, for $I_{li} \subseteq \{1, \ldots, n\}$ and $J_{lj} \subseteq \{1, \ldots, m\}$. And, since

$$\alpha_{li} \cap \beta_{lj} \doteq \sum_{i' \in I_{li}} \alpha_{i'} \cap \sum_{j' \in J_{lj}} \beta_{j'} \doteq \sum_{i' \in I_{li}, j' \in J_{lj}} (\alpha_{i'} \cap \beta_{j'}),$$

we conclude that $\pi(\alpha_0) \cap \pi(\beta_0) = \{\alpha_1 \cap \beta_1, \ldots, \alpha_1 \cap \beta_m, \ldots, \alpha_n \cap \beta_m\}$ is a support for $\alpha_0 \cap \beta_0$. □

Example 4. Given the regular expression $\alpha_1 = (b + ab + aab + abab) \cap (ab)^\star$, $\pi(\alpha_1) = \{bab \cap b(ab)^\star,\ ab \cap b(ab)^\star,\ b \cap b(ab)^\star,\ \varepsilon \cap b(ab)^\star,\ bab \cap (ab)^\star,\ ab \cap (ab)^\star,\ b \cap (ab)^\star,\ \varepsilon \cap (ab)^\star\}$.

The next proposition provides an upper bound on the cardinality of the support of a regular expression.

Proposition 5. *For all $\alpha \in \mathsf{RE}_\cap$, the inequality $|\pi(\alpha)| \leq 2^{|\alpha|_\Sigma - |\alpha|_\cap - 1}$ holds.*

Proof. We proceed by induction on the structure of the regular expression α. It is easily proved that the statement holds for the base cases ε, $\emptyset$ and $a \in \Sigma$. Assume that the result holds for some $\alpha, \beta \in \mathsf{RE}_\cap$. We will make use of the fact that $2^m + 2^n \leq 2^{m+n+1}$, for any $m, n \geq 0$. For $\alpha + \beta$, one has

$$\begin{aligned} |\pi(\alpha+\beta)| &= |\pi(\alpha) \cup \pi(\beta)| \leq |\pi(\alpha)| + |\pi(\beta)| \leq \\ &\leq 2^{|\alpha|_\Sigma - |\alpha|_\cap - 1} + 2^{|\beta|_\Sigma - |\beta|_\cap - 1} \leq \\ &\leq 2^{|\alpha|_\Sigma - |\alpha|_\cap - 1 + |\beta|_\Sigma - |\beta|_\cap - 1 + 1} = 2^{|\alpha+\beta|_\Sigma - |\alpha+\beta|_\cap - 1}. \end{aligned}$$

The case for $\alpha\beta$ is analogous. For $\alpha^\star$, one has

$$|\pi(\alpha^\star)| = |\pi(\alpha)\alpha^\star| = |\pi(\alpha)| \leq 2^{|\alpha|_\Sigma - |\alpha|_\cap - 1} = 2^{|\alpha^\star|_\Sigma - |\alpha^\star|_\cap - 1}.$$

Finally, for $\alpha \cap \beta$, one has

$$\begin{aligned} |\pi(\alpha \cap \beta)| &= |\pi(\alpha) \cap \pi(\beta)| \leq \\ &\leq |\pi(\alpha)| \cdot |\pi(\beta)| \leq 2^{|\alpha|_\Sigma - |\alpha|_\cap - 1} \cdot 2^{|\beta|_\Sigma - |\beta|_\cap - 1} = \\ &= 2^{|\alpha|_\Sigma - |\alpha|_\cap - 1 + |\beta|_\Sigma - |\beta|_\cap - 1} = 2^{|\alpha\cap\beta|_\Sigma - (|\alpha\cap\beta|_\cap - 1) - 2} = \\ &= 2^{|\alpha\cap\beta|_\Sigma - |\alpha\cap\beta|_\cap - 1}. \end{aligned}$$

□

The next examples present families of regular expressions that witnesses the tightness of the upper bound established in Proposition 5.

Example 6. Let the regular expression $r_n \in \mathsf{RE}_\cap$ over $\Sigma = \{a, b\}$ be inductively defined by $r_0 = a^\star b^\star$, $r_1 = b^\star a$ and $r_n = r_{n-2} \cap r_{n-1}^\star$, for $n \geq 2$. Using the definition of support it is straightforward that $|\pi(r_0)| = |\{a^\star b^\star, b^\star\}| = 2^1$, $|\pi(r_1)| = |\{b^\star a, \varepsilon\}| = 2^1$, and $|\pi(r_n)| = |\pi(r_{n-2})| \cdot |\pi(r_{n-1})|$, for $n \geq 2$. Thus, we obtain $|\pi(r_n)| = 2^{\mathsf{fib}(n)}$, for $n \geq 0$, and where $\mathsf{fib}(n)$ is the Fibonacci sequence. Also, $|r_0|_\Sigma - |r_0|_\cap - 1 = 2 - 0 - 1 = 1$, $|r_1|_\Sigma - |r_1|_\cap - 1 = 2 - 0 - 1 = 1$, and $|r_n|_\Sigma - |r_n|_\cap - 1 = |r_{n-2}|_\Sigma + |r_{n-1}|_\Sigma - |r_{n-2}|_\cap - |r_{n-1}|_\Sigma - 1 - 1 = (|r_{n-2}|_\Sigma - |r_{n-2}|_\cap - 1) + (|r_{n-1}|_\Sigma - |r_{n-1}|_\cap - 1)$, for $n \geq 2$. Consequently, $|r_n|_\Sigma - |r_n|_\cap - 1 = \mathsf{fib}(n)$, for $n \geq 0$. We conclude that $|\pi(r_n)| = 2^{|r_n|_\Sigma - |r_n|_\cap - 1}$, for $n \geq 0$.

Example 7. Let the regular expression $r_n \in \mathsf{RE}_\cap$ over $\{a\}$, be defined inductively by $r_0 = a^\star a$ and $r_n = r_{n-1} \cap a^\star a$, for $n \geq 1$. We have $\pi(r_0) = \pi(a^\star a) = \{a^\star a, \varepsilon\}$, and for $n \geq 1$,

$$\pi(r_n) = \underbrace{\{a^* a, \varepsilon\} \pcap \cdots \pcap \{a^* a, \varepsilon\}}_{n+1}.$$

Thus $|\pi(r_0)| = 2$ and $|\pi(r_n)| = |\pi(r_0)|^{n+1} = 2^{n+1}$. Note that $|r_n|_\Sigma = 2n+2$ and $|r_n|_\cap = n$. Therefore $|\pi(r_n)| = 2^{n+1} = 2^{2n+2-n-1} = 2^{|r_n|_\Sigma - |r_n|_\cap - 1}$.

4 Partial Derivatives

The notions of partial derivatives and partial derivative automata were introduced by Antimirov [1] for standard regular expressions. We now consider the Antimirov construction from $\mathsf{RE}_\cap$ expressions to NFAs.

Definition 8. *For a regular expression $\alpha \in \mathsf{RE}_\cap$ and a symbol $a \in \Sigma$, the set $\partial_a(\alpha)$ of* partial derivatives *of α w.r.t. a is defined by:*

$$\partial_a(\emptyset) = \emptyset, \qquad \partial_a(\alpha\beta) = \begin{cases} \partial_a(\alpha)\beta \cup \partial_a(\beta), & \text{if } \varepsilon(\alpha) = \varepsilon \\ \partial_a(\alpha)\beta & \text{otherwise,} \end{cases}$$
$$\partial_a(\varepsilon) = \emptyset,$$
$$\partial_a(b) = \begin{cases} \{\varepsilon\}, & \text{if } a = b \\ \emptyset & \text{otherwise,} \end{cases} \qquad \begin{aligned} \partial_a(\alpha + \beta) &= \partial_a(\alpha) \cup \partial_a(\beta), \\ \partial_a(\alpha \cap \beta) &= \partial_a(\alpha) \pcap \partial_a(\beta), \\ \partial_a(\alpha^\star) &= \partial_a(\alpha)\alpha^\star. \end{aligned}$$

This definition is extended to words $w \in \Sigma^\star$ 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 $R \subseteq \mathsf{RE}_\cap$. It follows easily that $\mathcal{L}(\partial_w(\alpha)) = w^{-1}\mathcal{L}(\alpha)$. The set of partial derivatives of an expression α is $\partial(\alpha) = \bigcup_{w \in \Sigma^\star} \partial_w(\alpha)$. We also define $\partial^+(\alpha) = \bigcup_{w \in \Sigma^+} \partial_w(\alpha)$.

As for standard regular expressions, the partial derivative automaton of an expression $\alpha \in \mathsf{RE}_\cap$ is defined by $\mathcal{A}_{pd}(\alpha) = \langle \partial(\alpha), \Sigma, \{\alpha\}, \delta_\alpha, F_\alpha \rangle$, where $F_\alpha = \{\ \gamma \in \partial(\alpha) \mid \varepsilon(\gamma) = \varepsilon\ \}$ and $\delta_\alpha(\gamma, a) = \partial_a(\gamma)$. It follows that $\mathcal{L}(\mathcal{A}_{pd}(\alpha))$ is exactly $\mathcal{L}(\alpha)$. Mirkin's and Antimirov's constructions coincide for standard regular expressions. We will see that this is not true for regular expressions with intersection.

The following lemmas present some properties of the function ∂_w, used to prove Proposition 11 and are easy to prove.

Lemma 9. *For all $S, S' \subseteq \mathsf{RE}_\cap$ and $a \in \Sigma$, the following property holds*

$$\partial_a(S \pcap S') = \partial_a(S) \pcap \partial_a(S').$$

Let $\mathrm{suff}(w)$ be the set of all non-empty suffixes of w, being defined as $\mathrm{suff}(w) = \{\ v \in \Sigma^+ \mid \exists u \in \Sigma^\star : uv = w\ \}$. Except for the second case, the following lemma was shown by Antimirov.

Lemma 10. *For every regular expressions $\alpha, \beta \in \mathsf{RE}_\cap$ and word $w \in \Sigma^+$, ∂_w satisfies the following:*

$$\partial_w(\alpha + \beta) = \partial_w(\alpha) \cup \partial_w(\beta), \tag{3}$$
$$\partial_w(\alpha \cap \beta) = \partial_w(\alpha) \cap \partial_w(\beta), \tag{4}$$
$$\partial_w(\alpha\beta) \subseteq \partial_w(\alpha)\beta \cup \bigcup_{v \in \mathrm{suff}(w)} \partial_v(\beta), \tag{5}$$
$$\partial_w(\alpha^\star) \subseteq \bigcup_{v \in \mathrm{suff}(w)} \partial_v(\alpha)\alpha^\star. \tag{6}$$

Proposition 11. *For every regular expressions $\alpha, \beta \in \mathsf{RE}_\cap$, the following holds.*

$$\partial^+(\alpha + \beta) \subseteq \partial^+(\alpha) \cup \partial^+(\beta), \qquad \partial^+(\alpha \cap \beta) \subseteq \partial^+(\alpha) \cap \partial^+(\beta),$$
$$\partial^+(\alpha\beta) \subseteq \partial^+(\alpha)\beta \cup \partial^+(\beta), \qquad \partial^+(\alpha^\star) \subseteq \partial^+(\alpha)\alpha^\star.$$

Proof. First note that, given a set $E \subseteq \mathsf{RE}_\cap$ and a regular expression $\alpha \in \mathsf{RE}_\cap$, if, for all $w \in \Sigma^+$, we have that $\partial_w(\alpha) \subseteq E$, then we have $\bigcup_{w \in \Sigma^+} \partial_w(\alpha) \subseteq E$ and thus $\partial^+(\alpha) \subseteq E$. Moreover, we know that for every $w \in \Sigma^+$, $\partial_w(\alpha) \subseteq \partial^+(\alpha)$, since $\partial^+(\alpha) = \bigcup_{w \in \Sigma^+} \partial_w(\alpha)$. Let $\alpha, \beta \in \mathsf{RE}_\cap$ be regular expressions over Σ. In order to prove the inclusions above, the facts mentioned above are used. The proof of each inclusion is given, respectively, by the following four proofs:

1. From equation (3), for all $w \in \Sigma^+$, the following holds:
$$\partial_w(\alpha + \beta) = \partial_w(\alpha) \cup \partial_w(\beta) \subseteq \partial^+(\alpha) \cup \partial^+(\beta).$$
And thus, we can conclude that $\partial^+(\alpha + \beta) \subseteq \partial^+(\alpha) \cup \partial^+(\beta)$.
2. In the same way, from equation (4), for all $w \in \Sigma^+$, the following holds:
$$\partial_w(\alpha \cap \beta) \subseteq \partial_w(\alpha) \cap \partial_w(\beta) \subseteq \partial^+(\alpha) \cap \partial^+(\beta).$$
And then, $\partial^+(\alpha \cap \beta) \subseteq \partial^+(\alpha) \cap \partial^+(\beta)$.
3. From equation (5), for all $w \in \Sigma^+$, the following holds:
$$\partial_w(\alpha\beta) \subseteq \partial_w(\alpha)\beta \cup \bigcup_{v \in \mathrm{suff}(w)} \partial_v(\beta) \subseteq \partial^+(\alpha)\beta \cup \partial^+(\beta).$$
Thus, $\partial^+(\alpha\beta) \subseteq \partial^+(\alpha)\beta \cup \partial^+(\beta)$.
4. Finally, from equation (6), for all $w \in \Sigma^+$, the following holds:
$$\partial_w(\alpha^\star) \subseteq \bigcup_{v \in \mathrm{suff}(w)} \partial_v(\alpha)\alpha^\star \subseteq \partial^+(\alpha)\alpha^\star.$$
Therefore, we have that $\partial^+(\alpha) \subseteq \partial^+(\alpha)\alpha^\star$. □

Example 12. Consider again $\alpha_1 = (b+ab+aab+abab) \cap (ab)^\star$. We have $\partial^+(\alpha_1) = \{bab \cap b(ab)^\star,\ ab \cap b(ab)^\star,\ b \cap b(ab)^\star,\ ab \cap (ab)^\star,\ \varepsilon \cap (ab)^\star\}$. Now, with $\beta = (b + ab + aab + abab)$, one has

$$\begin{aligned}\partial^+(\beta) \cap \partial^+((ab)^\star) = \{&bab \cap b(ab)^\star, ab \cap b(ab)^\star,\ b \cap b(ab)^\star,\\ &\varepsilon \cap b(ab)^\star,\ bab \cap (ab)^\star,\ ab \cap (ab)^\star,\ b \cap (ab)^\star,\ \varepsilon \cap (ab)^\star\}.\end{aligned}$$

Thus, we conclude that $\partial^+(\alpha_1) \subset \partial^+(b + ab + aab + abab) \cap \partial^+((ab)^\star)$.

The following proposition relates the function ∂^+ and the support π.

Proposition 13. *Given $\alpha \in RE_\cap$, $\partial^+(\alpha) \subseteq \pi(\alpha)$.*

Proof. The proof proceeds by induction on the structure of α. It is trivial that $\partial^+(\emptyset) = \pi(\emptyset)$, $\partial^+(\varepsilon) = \pi(\varepsilon)$ and $\partial^+(a) = \pi(a)$, for a symbol $a \in \Sigma$. Assume that $\partial^+(\alpha) \subseteq \pi(\alpha)$ and $\partial^+(\beta) \subseteq \pi(\beta)$ holds, for $\alpha, \beta \in \mathsf{RE}_\cap$. For $\alpha + \beta$, we have $\partial^+(\alpha+\beta) \subseteq \partial^+(\alpha) \cup \partial^+(\beta) \subseteq \pi(\alpha) \cup \pi(\beta)$. For $\alpha \cap \beta$, there is $\partial^+(\alpha \cap \beta) \subseteq \partial^+(\alpha) \cap \partial^+(\beta) \subseteq \pi(\alpha) \cap \pi(\beta)$. For $\alpha\beta$, we have $\partial^+(\alpha\beta) \subseteq \partial^+(\alpha)\beta \cup \partial^+(\beta) \subseteq \pi(\alpha)\beta \cup \pi(\beta)$. Finally, for $\alpha^\star$, $\partial^+(\alpha^\star) \subseteq \partial^+(\alpha)\alpha^\star \subseteq \pi(\alpha)\alpha^\star$. □

Since, for every regular expression $\alpha \in \mathsf{RE}_\cap$, the set $\pi(\alpha)$ is finite, Proposition 13 also proves that the set $\partial^+(\alpha)$ is finite. For regular expressions without intersection it is known that π and ∂^+ coincide [11]. Examples 4 and 12 show that there exists $\alpha \in \mathsf{RE}_\cap$ such that $\pi(\alpha) \neq \partial^+(\alpha)$. The following lemmas establish some conditions for the equality of $\pi(\alpha \cap \beta)$ and $\partial^+(\alpha \cap \beta)$ to hold for $\alpha, \beta \in \mathsf{RE}_\cap$, and will be used in Proposition 16.

Lemma 14. *Given $\alpha, \beta \in RE_\cap$, one has $\pi(\alpha \cap \beta) = \partial^+(\alpha \cap \beta)$ if and only if $\pi(\alpha) = \partial^+(\alpha)$, $\pi(\beta) = \partial^+(\beta)$ and $\partial^+(\alpha \cap \beta) = \partial^+(\alpha) \cap \partial^+(\beta)$.*

Proof. ($\Rightarrow$) We have that $\pi(\alpha \cap \beta) = \partial^+(\alpha \cap \beta) \subseteq \partial^+(\alpha) \cap \partial^+(\beta)$. From Proposition 13 follows that $\partial^+(\alpha) \subseteq \pi(\alpha)$ and $\partial^+(\beta) \subseteq \pi(\beta)$. Suppose by contradiction that $\partial^+(\alpha) \subset \pi(\alpha)$ or $\partial^+(\beta) \subset \pi(\beta)$. Then $\partial^+(\alpha \cap \beta) \subseteq \partial^+(\alpha) \cap \partial^+(\beta) \subset \pi(\alpha) \cap \pi(\beta) = \pi(\alpha \cap \beta)$, a contradiction since $\pi(\alpha \cap \beta) = \partial^+(\alpha \cap \beta)$. Thus, we conclude that $\pi(\alpha) = \partial^+(\alpha)$ and $\pi(\beta) = \partial^+(\beta)$. Consequently, $\pi(\alpha \cap \beta) = \pi(\alpha) \cap \pi(\beta) = \partial^+(\alpha \cap \beta)$.

($\Leftarrow$) This follows trivially from the definition of support, i.e., $\pi(\alpha \cap \beta) = \pi(\alpha) \cap \pi(\beta)$, since $\pi(\alpha) = \partial^+(\alpha)$ and $\pi(\beta) = \partial^+(\beta)$. □

Lemma 15. *Given $\alpha, \beta \in RE_\cap$, such that $\partial_w(\alpha) = \pi(\alpha)$ or $\partial_w(\beta) = \pi(\beta)$ holds for all $w \in \Sigma^+$, then $\partial^+(\alpha \cap \beta) = \partial^+(\alpha) \cap \partial^+(\beta)$.*

Proof. First, note that if $\gamma \in \mathsf{RE}_\cap$ and $\partial_w(\gamma) = \pi(\gamma)$ for every $w \in \Sigma^+$, then $\partial^+(\gamma) = \bigcup_{w \in \Sigma^+} \partial_w(\gamma) = \pi(\gamma)$. Given $\alpha, \beta \in \mathsf{RE}_\cap$, there are three possible cases to prove. First, suppose that, for all $w \in \Sigma^+$, we have $\partial_w(\alpha) = \pi(\alpha)$ and $\partial_w(\beta) = \pi(\beta)$. Then

$$\partial^+(\alpha \cap \beta) = \bigcup_{w \in \Sigma^+} (\partial_w(\alpha) \cap \partial_w(\beta)) = \pi(\alpha) \cap \pi(\beta) = \partial^+(\alpha) \cap \partial^+(\beta).$$

It remains to prove the cases that either $\partial_w(\alpha) = \pi(\alpha)$ or $\partial_w(\beta) = \pi(\beta)$, for all $w \in \Sigma^+$. The proof is the same for both cases. So, we will only present the proof

for the first case. Suppose that, for all $w \in \Sigma^+$, $\partial_w(\alpha) = \pi(\alpha)$, it holds that

$$\begin{aligned}
\partial^+(\alpha \cap \beta) &= \bigcup_{w\in\Sigma^+} (\partial_w(\alpha) \cap \partial_w(\beta) = \bigcup_{w\in\Sigma^+} (\pi(\alpha) \cap \partial_w(\beta)) \\
&= \bigcup_{w\in\Sigma^+} \{\alpha_i \cap \beta_j \mid \alpha_i \in \pi(\alpha),\ \beta_j \in \partial_w(\beta)\} \\
&= \left\{ \alpha_i \cap \beta_j \;\middle|\; \alpha_i \in \pi(\alpha),\ \beta_j \in \bigcup_{w\in\Sigma^+} \partial_w(\beta) \right\} \\
&= \{\alpha_i \cap \beta_j \mid \alpha_i \in \pi(\alpha),\ \beta_j \in \partial^+(\beta)\} \\
&= \pi(\alpha) \cap \partial^+(\beta) = \partial^+(\alpha) \cap \partial^+(\beta).
\end{aligned}$$

□

By Proposition 13, $|\pi(\alpha)|$ is an upper bound for the cardinality of $\partial^+(\alpha)$. This upper bound can be achieved, as shown by the following proposition.

Proposition 16. *For any $n \in \mathbb{N}$ there exists a regular expression $r_n \in \mathsf{RE}_\cap$ of size $O(n)$ such that $|\partial^+(r_n)| = 2^{|r_n|_\Sigma - |r_n|_\cap - 1}$.*

Proof. Consider the regular expressions $r_n \in \mathsf{RE}_\cap$ from Example 7. We prove that $\pi(r_n) = \partial^+(r_n)$. The proof proceeds by induction on n. For $n = 0$ and for all $w \in \Sigma^+$, we have $\partial_w(a^\star a) = \{a^\star a, \epsilon\} = \partial^+(a^\star a) = \pi(a^\star a)$. Let us assume, by induction, that $\pi(r_n) = \partial^+(r_n)$, for $n \geq 1$. It follows from Lemma 15 that $\partial^+(r_{n+1}) = \partial^+(r_n \cap a^\star a) = \partial^+(r_n) \cap \partial^+(a^\star a)$. Since $\pi(a^\star a) = \partial^+(a^\star a)$, $\pi(r_n) = \partial^+(r_n)$, and $\partial^+(r_n \cap a^\star a) = \partial^+(r_n) \cap \partial^+(r_n)$, we conclude, from Lemma 14, that $\pi(r_{n+1}) = \pi(r_n \cap a^\star a) = \partial^+(r_n \cap a^\star a) = \partial^+(r_{n+1})$. □

The next example provides another non-trivial family of regular expressions for which the set of partial derivatives and the support coincide.

Example 17. For $n \geq 0$ let the regular expression $s_n \in \mathsf{RE}_\cap$ be inductively defined by $s_0 = (a+b)^\star b(a+b)^\star$ and $s_n = ((a+b)s_{n-1}(a+b)) \cap ((a+b)^\star(a+b))$, for $n \geq 1$. The alphabetic length of s_n is $|s_n|_\Sigma = 5 + 8n$ and $|s_n|_\cap = n$. The cardinality of the support of s_n is given by: $|\pi(s_0)| = 2$, $|\pi(s_1)| = 6$ and $|\pi(s_n)| = \sum_{i=2}^{n} 2^i + 3 \cdot 2^n$, for $n \geq 2$ Thus, for $n \geq 2$ we have $|\pi(s_n)| = O(2^n)$. Let $m = |s_n|_\Sigma - |s_n|_\cap - 1 = 5 + 7n - 1$, i.e. $n = (m-4)/7$. Then, $|\pi(s_n)| = O(2^{\frac{1}{7}m}) = O(1.105^m)$, which is much smaller than the upper bound 2^m. For all $n \geq 0$, $\pi(s_n) = \partial^+(s_n)$.

5 Average Complexity Results

We know that the number of states in the partial derivative automaton of an expression α has $|\pi(\alpha)|$ as its tight upper bound. In this section we estimate an upper bound for the asymptotic average size of $\pi(\alpha)$. This is done using standard methods of analytic combinatorics as expounded by Flajolet and Sedgewick [13],

which apply to generating functions $f(z) = \sum_n a_n z^n$ associated with combinatorial classes. Given some measure of the objects of a combinatorial class $\mathcal{A}$, the coefficient a_n represents the sum of the values of this measure for all objects of size n. We will use the notation $[z^n]f(z)$ for a_n. For an introduction to this approach applied to formal languages, we refer to Broda *et al.* [6].

Although the methods here used are the standard ones from the Analytic Combinatorics (and Complex Analysis), each application of these techniques is always a challenge, as one cannot foresee the analytic difficulties that one can incur into when conducting the study of the generation function. The generating function f can be seen as a complex analytic function, and the study of its behaviour near its dominant singularity η (in case there is only one, as it happens with the functions here considered) gives us access to the asymptotic form of its coefficients. In particular, if $f(z)$ is analytic in some appropriate neighbourhood of 0 containing η, then one has the following [13,6]:

Proposition 18. *If* $f(z) = a - b\sqrt{1 - z/\rho} + o\left(\sqrt{1 - z/\rho}\right)$, *with* $a, b \in \mathbb{R}$, $b \neq 0$, *then*

$$[z^n]f(z) \sim \frac{b}{2\sqrt{\pi}}\, \rho^{-n} n^{-3/2}.$$

If $f(z) = \frac{a}{\sqrt{1-z/\rho}} + o\left(\frac{1}{\sqrt{1-z/\rho}}\right)$, *with* $a \in \mathbb{R}$, *and* $a \neq 0$, *then*

$$[z^n]f(z) \sim \frac{a}{\sqrt{\pi}}\, \rho^{-n} n^{-1/2}.$$

5.1 Number of Expressions and Letters and ∩ symbols

The study of the combinatorial behaviour of the $\mathsf{RE}_\cap$-expressions, both in terms of the number of expressions and the number of letters in them, is identical to the study of any other regular expressions with 3 binary operators and a single unary operator. Thus the results presented in Broda et al. [7] are valid for the case here studied. Denoting by $R_k(z)$ the generating function for the number of $\mathsf{RE}_\cap$-expressions without $\emptyset$ over a k letters alphabet, and by $L_k(z)$ the generating function for the number of letters in the expressions, one has:

$$[z^n]R_k(z) \sim c_k \rho_k^{-n-\frac{1}{2}} n^{-\frac{3}{2}}, \tag{7}$$

$$[z^n]L_k(z) \sim \frac{k}{12\pi c_k} \rho_k^{-n+\frac{1}{2}} n^{-\frac{1}{2}}, \tag{8}$$

where $c_k = \frac{\sqrt[4]{3+3k}}{6\sqrt{\pi}}$ and $\rho_k = \frac{-1+2\sqrt{3+3k}}{11+12k}$.

The average number of letters in an expression of size n is given by

$$\frac{[z^n]L_k(z)}{[z^n]R_k(z)}.$$

Using equations (7) and (8), one obtains, asymptotically,

$$|\alpha|_{\Sigma} \sim \frac{3k\rho_k}{\sqrt{3+3k}}\,\|\alpha\| \xrightarrow[k\to\infty]{} \frac{1}{2}\|\alpha\|. \tag{9}$$

The number of intersections in the $\mathsf{RE}_{\cap}$-expressions under consideration can be computed as follows. Consider the bivariate generating function

$$\mathcal{I}_k(u,z) = \sum_{m,n} \iota_{mn} u^m z^n,$$

where ι_{mn} is the number of $\mathsf{RE}_{\cap}$-expressions with m intersection symbols and size n. From (1), and using the symbolic method, we can write

$$\mathcal{I}_k(u,z) = (k+1)z + 2z\mathcal{I}_k(u,z)^2 + uz\mathcal{I}_k(u,z)^2 + z\mathcal{I}_k(u,z).$$

Solving this for $\mathcal{I}_k(u,z)$, differentiating the result w.r.t. u, and making $u=1$, we obtain an expression for the generating function for the cumulative number of intersection symbols in all $\mathsf{RE}_{\cap}$-expressions of size n:

$$I_k(z) = \frac{1}{18z}\sqrt{q_k(z)} + \frac{(k+1)z}{3\sqrt{q_k(z)}} + \frac{z-1}{18z}, \tag{10}$$

where $q_k(z) = 1 - 2z - (11+12k)z^2$, from which one obtains, using the same methods,

$$[z^n]I_k(z) \sim \frac{1}{6\sqrt{\pi}}\left(\frac{(k+1)\sqrt{\rho_k}}{\sqrt[4]{3+3k}\sqrt{n}} - \frac{\sqrt[4]{3+3k}}{3\sqrt{\rho_k}\,n^{3/2}}\right)\rho_k^{-n}. \tag{11}$$

The average number of symbols $\cap$ in an expression of size n is given by

$$\frac{[z^n]I_k(z)}{[z^n]R_k(z)}.$$

Using equations (7) and (11), one obtains, asymptotically,

$$|\alpha|_{\cap} \sim \frac{(k+1)\rho_k}{\sqrt{3+3k}}\,\|\alpha\| \xrightarrow[k\to\infty]{} \frac{1}{6}\|\alpha\|. \tag{12}$$

5.2 Average Size of π

Let $P_k(z)$ denote the generating function for the size of $\pi(\alpha)$ for expressions without $\emptyset$. From Definition 2 it follows that, given an expression α, an upper bound, $p(\alpha)$, for the number of elements[3] in the set $\pi(\alpha)$ satisfies:

$$\begin{aligned} p(\varepsilon) &= 0, & p(\alpha+\beta) &= p(\alpha)+p(\beta),\\ p(a) &= 1, \quad \text{for } a\in\Sigma, & p(\alpha\beta) &= p(\alpha)+p(\beta),\\ p(\alpha^\star) &= p(\alpha), & p(\alpha\cap\beta) &= p(\alpha)p(\beta). \end{aligned}$$

[3] This upper bound corresponds to the case where all unions in $\pi(\alpha)$ are disjoint.

From this, we directly get

$$P_k(z) = kz + 4zP_k(z)R_k(z) + zP_k(z) + zP_k(z)^2,$$

from which we obtain the following closed expression

$$P_k(z) = \frac{1 - z + 2\sqrt{q_k(z)} - \sqrt{p_k(z) + 4(1-z)\sqrt{q_k(z)}}}{6z}, \tag{13}$$

where

$$p_k(z) = 5 - 10z - (43 + 84k)z^2. \tag{14}$$

One now needs to determine the dominant singularity of $P_k(z)$ which can either be a root of $q_k(z)$ or a root of $r_k(z) = p_k(z) + 4(1-z)\sqrt{q_k(z)}$. We need to know which of the two expressions $r_k(z)$ or $q_k(z)$ has the smallest positive zero. Because this is not trivial (note that one needs to decide this for all k), one will do it indirectly using the method expounded in the following paragraphs.

Observing that $r_k(0) = 9$ is positive and

$$r_k(\rho_k) = \frac{12\left(13 - 14k - 24k^2 + (8k-4)\sqrt{3+3k}\right)}{(11+12k)^2} < 0,$$

by Bolzano theorem, $r_k(z)$ must have a positive zero smaller than ρ_k. This conclusion could be achieved, directly, from the fact that the absolute value of the negative zero of $q_k(z)$ is smaller than its positive zero, and thus, by Pringsheim theorem [13], another smaller positive singularity of $P_k(z)$ necessarily exists that can only be due to $r_k(z)$. Letting

$$\bar{\rho}_k = \frac{-1 - 2\sqrt{3+3k}}{11+12k},$$

and observing that

$$r_k(\bar{\rho}_k) = -\frac{12\left(-13 + 14k + 24k^2 + (8k-4)\sqrt{3+3k}\right)}{(11+12k)^2} < 0,$$

one concludes that $r_k(z)$ has necessarily two real zeros in its domain, $[\bar{\rho}_k, \rho_k]$. Analogously, $s_k(z) = p_k(z) - 4(1-z)\sqrt{q_k(z)}$ has also two real zeros in the same interval, and since $r_k(z)s_k(z)$ is a fourth degree polynomial, it follows that $r_k(z)$ has exactly two zeros, η_k and η_k', which are real. Since $s_k(0) = 1 < r_k(0) = 9$, and $r_k(x) = s_k(x)$ only at the end points of $[\bar{\rho}_k, \rho_k]$ it follows that $s_k(x) < r_k(x)$ in $]\bar{\rho}_k, \rho_k[$. Considering the four real zeros of the polynomial $r_k(z)s_k(z)$, given what we just said, we conclude that the two more distant zeros from the origin are the roots of $r_k(z)$. In fact, we can obtain an explicit expression for the zeros of $r_k(z)s_k(z)$ by noticing that

$$\begin{aligned} p_k(z) \pm 4(1-z)\sqrt{q_k(z)} &= \left(1 - z \pm 2\sqrt{q_k(z)}\right)^2 - 36kz^2 \\ &= \left(1 - z \pm 2\sqrt{q_k(z)} - 6\sqrt{k}z\right)\left(1 - z \pm 2\sqrt{q_k(z)} + 6\sqrt{k}z\right), \end{aligned}$$

and thus, solving the equations resulting of nulling those factors, we obtain the four zeros of $r_k(z)s_k(z)$:

$$\begin{aligned} \eta_k &= \frac{4\sqrt{2k+1}+2\sqrt{k}-1}{28k+4\sqrt{k}+15}, \qquad \eta'_k = -\frac{4\sqrt{2k+1}+2\sqrt{k}+1}{28k-4\sqrt{k}+15}, \\ \eta''_k &= \frac{4\sqrt{2k+1}-2\sqrt{k}-1}{28k-4\sqrt{k}+15}, \qquad \eta'''_k = -\frac{4\sqrt{2k+1}-2\sqrt{k}+1}{28k+4\sqrt{k}+15}. \end{aligned} \tag{15}$$

It is possible to verify that η_k and η'_k are the roots of $r_k(z)$ and the other two the roots from $s_k(z)$. Therefore, one has

$$r_k(z)s_k(z) = (7056k^2+7416k+2025)(z-\eta_k)(z-\eta'_k)(z-\eta''_k)(z-\eta'''_k). \tag{16}$$

From (13) one has

$$6zP_k(z) = 1 - z - \sqrt{r_k(z)} + 2\sqrt{q_k(z)}, \tag{17}$$

and we split the study of the coefficients of the series of $P_k(z)$ into the study of the coefficients of $1-z-\sqrt{r_k(z)}$ and of $2\sqrt{q_k(z)}$. For the first one, we use that

$$r_k(z) = \frac{7056k^2+7416k+2025}{s_k(z)}\eta_k(\eta'_k-z)(\eta''_k-z)(\eta'''_k-z)\left(1-\frac{z}{\eta_k}\right),$$

and the fact that given a complex function f, defined in a neighbourhood of η such that $\lim_{z\to\eta} f(z) = a$, one has, for all $r \in \mathbb{R}$, $f(z)(1-z/\eta)^r = a(1-z/\eta)^r + o((1-z/\eta)^r)$, together with Proposition 18, to obtain

$$[z^n]\left(1-z-\sqrt{r_k(z)}\right) \sim \lambda_k \eta_k^{-n} n^{-\frac{3}{2}},$$

where

$$\lambda_k = \left(\frac{(7056k^2+7416k+2025)(\eta'_k-\eta_k)(\eta''_k-\eta_k)(\eta'''_k-\eta_k)\eta_k}{2\pi s_k(\eta_k)}\right)^{\frac{1}{2}}. \tag{18}$$

For the last summand one has, similarly,

$$2\sqrt{q_k(z)} = 4\sqrt[4]{3+3k}\,\rho_k^{\frac{1}{2}}(\rho_k-\bar{\rho}_k)^{\frac{1}{2}}(1-z/\rho_k)^{\frac{1}{2}} + o\left((1-z/\rho_k)^{\frac{1}{2}}\right),$$

from which it follows, $[z^n]2\sqrt{q_k(z)} \sim -\mu_k\rho_k^{-n}n^{-\frac{3}{2}}$, where

$$\mu_k = 2\pi^{-\frac{1}{2}}\rho_k^{\frac{1}{2}}\sqrt[4]{3+3k}. \tag{19}$$

Summing up, we get that

$$[z^n]P_k(z) \sim \frac{1}{6}\left(\lambda_k\eta_k^{-(n+1)} - \mu_k\rho_k^{-(n+1)}\right)n^{-\frac{3}{2}}. \tag{20}$$

In order to see what this result entails for the average case when compared with the worst case result, expressed in Proposition 5, attend to the following.

$$\left(\frac{[z^n]P_k(z)}{[z^n]R_k(z)}\right)^{\frac{1}{n}} \sim \left(\frac{\frac{1}{6}\lambda_k\eta_k^{-(n+1)}n^{-\frac{3}{2}}}{c_k\rho_k^{-n-\frac{1}{2}}(n+1)^{-\frac{3}{2}}}\right)^{\frac{1}{n}} \xrightarrow[n\to\infty]{} \frac{\rho_k}{\eta_k}.$$

Setting $\gamma_k = \frac{\rho_k}{\eta_k}$, this means that, on average,

$$|\pi(\alpha)| \sim \gamma_k^{\|\alpha\|}.$$

One has $\gamma_2 \sim 1.01655$, $\gamma_{10} \sim 1.04137$, $\gamma_{100} \sim 1.05294$, and

$$\lim_{k\to\infty} \gamma_k = \frac{7\sqrt{3}}{6\sqrt{2}+3} \sim 1.05564.$$

Proposition 19. *For large values of k and n an upper bound for the average number of states of $\mathcal{A}_{pd}$ is $(1.056 + o(1))^n$.*

Considering the estimates given in (9) and (12), the worst-case upper bound $2^{|\alpha|_\Sigma - |\alpha|_\cap - 1}$ from Proposition 5 leads to an upper bound for the average case roughly of $\sqrt[3]{2}^{\|\alpha\|}$, for α large enough. As $\sqrt[3]{2} \sim 1.25992$, the result just obtained shows that the upper bound for the average complexity is significantly smaller than the one for the worst case.

6 Conclusions

The conversion of a regular expression with intersection α to NFA is in the worst-case $2^{\Omega(\|\alpha\|)}$ [15,18,17]. This fact leads to the assumption that, although succinct, these expressions are not useful in practical applications. Here we show that, asymptotically, an upper bound for the average-state complexity of $\mathcal{A}_{pd}(\alpha)$ is exponential but with a base only slightly above 1. Actually, experimental results using a uniform distribution suggest that the average-state complexity of $\mathcal{A}_{pd}(\alpha)$ may even be polynomial [3].

References

1. Antimirov, V.: Partial derivatives of regular expressions and finite automaton constructions. Theoret. Comput. Sci. 155(2), 291–319 (1996)
2. Antimirov, V.M., Mosses, P.D.: Rewriting extended regular expressions. In: Rozenberg, G., Salomaa, A. (eds.) 1st DLT. pp. 195–209. World Scientific (1994)
3. Bastos, R.: Manipulation of Extended Regular Expressions with Derivatives. Master's thesis, Faculdade de Ciências da Universidade do Porto (2015)
4. Broda, S., Machiavelo, A., Moreira, N., Reis, R.: On the average state complexity of partial derivative automata. Int. J. Found. Comput. S. 22(7), 1593–1606 (2011)
5. 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)
6. Broda, S., Machiavelo, A., Moreira, N., Reis, R.: A hitchhiker's guide to descriptional complexity through analytic combinatorics. Theor. Comput. Sci. 528, 85–100 (2014)
7. Broda, S., Machiavelo, A., Moreira, N., Reis, R.: Partial derivative automaton for regular expressions with shuffle. In: Shallit, J., Okhotin, A. (eds.) 17th DCFS. pp. 21–32. No. 9118 in LNCS, Springer (2015)
8. Brzozowski, J.A.: Derivatives of regular expressions. JACM 11(4), 481–494 (1964)

9. Caron, P., Champarnaud, J., Mignot, L.: Partial derivatives of an extended regular expression. In: Dediu, A.H., Inenaga, S., Martín-Vide, C. (eds.) 5th LATA. LNCS, vol. 6638, pp. 179–191. Springer (2011)
10. Caron, P., Champarnaud, J., Mignot, L.: A general framework for the derivation of regular expressions. RAIRO - Theor. Inf. and Applic. 48(3), 281–305 (2014)
11. Champarnaud, J.M., Ziadi, D.: From Mirkin's prebases to Antimirov's word partial derivatives. Fundam. Inform. 45(3), 195–205 (2001)
12. Christiansen, T., brian d foy, Wall, L., Orwant, J.: Programming Perl. O'Reilly Media (2012), 4th edition
13. Flajolet, P., Sedgewick, R.: Analytic Combinatorics. CUP (2008)
14. Fürer, M.: The complexity of the inequivalence problem for regular expressions with intersection. In: de Bakker, J.W., van Leeuwen, J. (eds.) 7th ICALP. LNCS, vol. 85, pp. 234–245. Springer (1980)
15. Gelade, W.: Succinctness of regular expressions with interleaving, intersection and counting. Theor. Comput. Sci. 411(31-33), 2987–2998 (2010)
16. Gelade, W., Neven, F.: Succinctness of the complement and intersection of regular expressions. In: Albers, S., Weil, P. (eds.) 25th STACS. LIPIcs, vol. 1, pp. 325–336. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany (2008)
17. Gruber, H.: On the descriptional and algorithmic complexity of regular languages. Ph.D. thesis, Justus Liebig University Giessen (2010)
18. Gruber, H., Holzer, M.: Finite automata, digraph connectivity, and regular expression size. In: Aceto, L., Damgård, I., Goldberg, L.A., Halldórsson, M.M., Ingólfsdóttir, A., Walukiewicz, I. (eds.) 35th ICALP. LNCS, vol. 5126, pp. 39–50. Springer (2008)
19. Jiang, T., Ravikumar, B.: A note on the space complexity of some decision problems for finite automata. Information Processing Letters 40(1), 25–31 (1991)
20. Mirkin, B.G.: An algorithm for constructing a base in a language of regular expressions. Engineering Cybernetics 5, 51–57 (1966)
21. Petersen, H.: The membership problem for regular expressions with intersection is complete in logcfl. In: Alt, H., Ferreira, A. (eds.) 19th STACS. LNCS, vol. 2285, p. 513–522. Springer (2002)
22. Sen, K., Rosu, G.: Generating optimal monitors for extended regular expressions. Electr. Notes Theor. Comput. Sci. (2003)
23. van der Vlist, E.: RELAX NG. O'Reilly Media (2003)