CORRIGENDUM TO OUR PAPER: HOW EXPRESSIONS CAN CODE FOR AUTOMATA

Sylvain Lombardy[1] and Jacques Sakarovitch[2]

Abstract. We correct a mistake made in a previous paper in the construction of an automaton from a rational expression. We used there the definition of derivation of expression given by Antimirov, and this definition has to be further adapted for our purpose.

1991 Mathematics Subject Classification. 68Q45, 68Q70.

A disturbing example

In [7], we were considering the following problem: *Is it possible to build an algorithm* Ω *such that for any rational expression* E *computed from an automaton* $\mathcal{A}$ — i.e. $\mathsf{E} = \Phi(\mathcal{A})$ *where* Φ *is the state elimination method for instance — the following holds:* $\mathcal{A} = \Omega(\mathsf{E})$ *?* We did not solve the problem completely, but we have identified two constructions that are good candidates to be the core components of such an algorithm Ω. The first one is the construction of an automaton $\Delta(\mathsf{E})$ from an expression E, the second one is the computation of the *minimal co-quotient* $\Upsilon(\mathcal{B})$ of an automaton $\mathcal{B}$. There is no problem with the minimal co-quotient but the definition we gave for $\Delta(\mathsf{E})$ was faulty. This can be observed for instance on the following example.

Let $\mathcal{A}_1$ be the automaton of Figure 1 (a), and let $\mathsf{E}_1 = (a + b + 1)\,[a(a + b)]^*$ be the expression computed from $\mathcal{A}_1$, which we write $\mathsf{E}_1 = (a + b + 1)\mathsf{F}_1$ with $\mathsf{F}_1 = [a(a + b)]^*$; Figure 1 (b) shows $\Delta(\mathsf{E}_1)$, whose co-quotient is not isomorphic to $\mathcal{A}_1$.

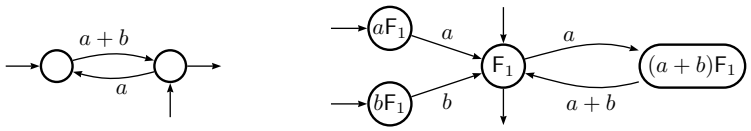
As we shall see, it is not difficult to recover correct definitions and a true statement (Theorem 1.6) for the key result (Theorem 3.5) in the original paper.

Keywords and phrases: finite automata, regular expression, derivation of expressions, quotient of automata.

[1] IGM-LabInfo (UMR 8049), Université Paris-Est Marne-la-Vallée, 77454 Marne-la-Vallée Cedex 2, France; e-mail: `lombardy@univ-mlv.fr`.

[2] LTCI (UMR 5141), CNRS / ENST, 46 rue Barrault, 75634 Paris Cedex 13, France; e-mail: `sakarovitch@enst.fr`.

What proves to be more difficult, or at least much longer, is the complete and detailled proof, hopefully correct now, of Theorem 1.6.



(a) $\mathcal{A}_1$ (b) $\Delta(\Phi(\mathcal{A}_1))$

FIGURE 1. A counter-example to the statement in [7]

1. The corrected definitions and statement

In order to make this corrigendum as self-contained as possible, let us first recall the notion of derivation as defined by Antimirov. If E is a rational expression on A, we write $|\mathsf{E}|$ for the language of A^* denoted by E and $\mathsf{c}(\mathsf{E})$ for the Boolean whose value is 1 if $|\mathsf{E}|$ contains the empty word and 0 otherwise; $\mathsf{c}(\mathsf{E})$ is effectively computed by induction on the depth of E.

Definition 1.1 (**Antimirov** [2])**.** Let E be a rational expression on A and let a be a letter in A. The $\mathbb{B}$-*derivative*[1] of E with respect to a, denoted $\dfrac{\partial}{\partial a}\mathsf{E}$, is *a set of rational expressions on* A, recursively defined by

$$\frac{\partial}{\partial a}\mathsf{0} = \frac{\partial}{\partial a}\mathsf{1} = \emptyset \ ,$$

$$\forall a, b \in A \qquad \frac{\partial}{\partial a} b = \begin{cases} \{\mathsf{1}\} & \text{if} \quad b = a \ , \\ \emptyset & \text{otherwise,} \end{cases}$$

$$\frac{\partial}{\partial a}(\mathsf{E}{+}\mathsf{F}) = \frac{\partial}{\partial a}\mathsf{E} \cup \frac{\partial}{\partial a}\mathsf{F} \ , \tag{1}$$

$$\frac{\partial}{\partial a}(\mathsf{E}\cdot\mathsf{F}) = \left[\frac{\partial}{\partial a}\mathsf{E}\right]\cdot\mathsf{F} \cup \mathsf{c}(\mathsf{E})\ \frac{\partial}{\partial a}\mathsf{F} \ , \tag{2}$$

$$\frac{\partial}{\partial a}(\mathsf{E}^*) = \left[\frac{\partial}{\partial a}\mathsf{E}\right]\cdot\mathsf{E}^* \ . \tag{3}$$

The induction implied by $(1-3)$ should be interpreted while distributing derivation and product over union:

$$\frac{\partial}{\partial a}\left[\bigcup_{i\in I}\mathsf{E}_i\right] = \bigcup_{i\in I}\frac{\partial}{\partial a}\mathsf{E}_i \ , \qquad \left[\bigcup_{i\in I}\mathsf{E}_i\right]\cdot\mathsf{F} = \bigcup_{i\in I}(\mathsf{E}_i\cdot\mathsf{F}) \ .$$

[1]We call it "$\mathbb{B}$-derivative" and not simply "derivative" for two reasons. First in order to avoid confusion with the derivation defined by Brzozowski [3], and second because the formulae depend on the semiring of multiplicities and can be defined for other semirings (*cf.* [6]).

Definition 1.2. Let E be a rational expression on A and g a non empty word of A^*, *i.e.* $g = f\,a$ with a in A. The $\mathbb{B}$-*derivative* of E with respect to g, denoted $\dfrac{\partial}{\partial g}\mathsf{E}$, is *the set of rational expressions over* A, recursively defined by formulae (1) – (3) and by

$$\forall f \in A^+\,,\ \forall a \in A \qquad \frac{\partial}{\partial fa}\mathsf{E} = \frac{\partial}{\partial a}\left(\frac{\partial}{\partial f}\mathsf{E}\right)\quad . \tag{4}$$

We call *true derived term* of E any rational expression which belongs to a set $\dfrac{\partial}{\partial g}\mathsf{E}$ for some g in A^+ and *derived term* the expression E itself, or any true derived term; we write $\mathrm{D}(\mathsf{E})$ for the set of derived terms of E.

It will be useful to distinguish the set of derived terms that are obtained as the result of a derivation by a word w: $\mathrm{D}_w(\mathsf{E}) = \{\mathsf{K} \mid \exists u \in A^* \quad \mathsf{K} \in \dfrac{\partial}{\partial uw}\mathsf{E}\}$.

Definition 1.3. The *derived term automaton* of a rational expression E is the finite automaton $\mathrm{A}(\mathsf{E})$ whose set of states is $\mathrm{D}(\mathsf{E})$ and whose transitions are defined by:
(i) if K and K' are derived terms of E and if a is a letter of A, $(\mathsf{K}, a, \mathsf{K}')$ is a transition if and only if K' belongs to $\frac{\partial}{\partial a}\mathsf{K}$;
(ii) the initial state is E;
(iii) a derived term K is final if and only if $\mathsf{c}(\mathsf{K}) = 1$.

The essence of the derivation (by a letter), as defined by Antimirov and repeated here, is to "break" the expression into pieces when the operator at the upper level of the expression is "+". The modification of the derivation we consider now consists in supposing that this breaking happens *spontaneously*, as if it were a derivation with respect to the empty word, before the first derivation by a letter, and after every such derivation. To that end, we define a new operation on rational expressions which we denote by d() and which, roughly speaking, consists of decomposing an expression into a set of expressions whose left factor is not a sum.

Here lie the correction we make to our original paper. The definition of the derivation we apply to the expressions is modified in two ways: first, the breaking d() is more 'complete' as it 'goes through' sets of expressions which contain 1; second, we replace the derivation by the breaking derivation (in the original paper, the breaking d() was applied only once before any derivation).

The definition of d() requires two further notations. If X is a *set of expressions*, then $[X]_{\mathsf{p}}$ is the same set possibly without the expression 1: $[X]_{\mathsf{p}} = X \setminus \{1\}$, and δ_X is the Boolean that takes the value 1 if the expression 1 belongs to X and 0 otherwise. For instance, $\delta_{[X]_{\mathsf{p}}} = 0$ for any set X. If E is an expression, $\mathrm{d}(\mathsf{E})$ is a set of expressions inductively defined by:

$$\mathrm{d}(0) = \{0\}\,, \qquad \mathrm{d}(1) = \{1\}\,, \qquad \forall a \in A \quad \mathrm{d}(a) = \{a\},$$
$$\mathrm{d}(\mathsf{E}+\mathsf{F}) = \mathrm{d}(\mathsf{E}) \cup \mathrm{d}(\mathsf{F})\ , \quad \mathrm{d}(\mathsf{E}\cdot\mathsf{F}) = [\mathrm{d}(\mathsf{E})]_{\mathsf{p}} \cdot \mathsf{F} \cup \delta_{\mathrm{d}(\mathsf{E})}\,\mathrm{d}(\mathsf{F})\ , \quad \mathrm{d}(\mathsf{E}^*) = \{\mathsf{E}^*\}\,,$$

and $\mathrm{d}(\mathsf{E})$ is called the set of *initial broken (derived) terms* of E.

Definition 1.4. The *breaking derivation* of an expression E with respect to a letter a is defined as:

$$\frac{\partial_{\mathsf{b}}}{\partial a}\mathsf{E} = \mathrm{d}\left(\left[\frac{\partial}{\partial a}\mathsf{E}\right]\right).$$

This breaking derivation is then extended to words by composition as in (4) and we call *true broken derived term* of E any rational expression which belongs to a set $\dfrac{\partial_{\mathsf{b}}}{\partial g}\mathsf{E}$ for some non empty word g in A^+:

$$\mathrm{TBD}(\mathsf{E}) = \{\mathsf{L} \mid \exists g \in A^+ \quad \mathsf{L} \in \frac{\partial_{\mathsf{b}}}{\partial g}\mathsf{E}\} \ .$$

We call *broken derived term* of E any element of the union of $\mathrm{d}(\mathsf{E})$ and $\mathrm{TBD}(\mathsf{E})$:

$$\mathrm{BD}(\mathsf{E}) = \mathrm{d}(\mathsf{E}) \cup \mathrm{TBD}(\mathsf{E}) \ .$$

It is easy to check that

$$\forall f \in A^* \qquad \frac{\partial_{\mathsf{b}}}{\partial f}\mathsf{E} = \mathrm{d}\left(\left[\frac{\partial}{\partial f}\mathsf{E}\right]\right) \ ,$$

which amounts to say that the broken derived terms are obtained by "breaking" the derived terms:

$$\mathrm{BD}(\mathsf{E}) = \bigcup_{\mathsf{K}\in \mathrm{D}(\mathsf{E})} \mathrm{d}(\mathsf{K}) \ . \tag{5}$$

The *broken derived term automaton* $\Delta'(\mathsf{E})$ is then defined as was the derived term automaton by Antimirov. Its set of states is $\mathrm{BD}(\mathsf{E})$, the set of initial states is $\mathrm{d}(\mathsf{E})$, a state K is final if and only if $\mathsf{c}(\mathsf{K}) = 1$ and for a in A, and K, K' in $\mathrm{BD}(\mathsf{E})$, $(\mathsf{K}, a, \mathsf{K}')$ is a transition if and only if K' belongs to $\dfrac{\partial_{\mathsf{b}}}{\partial a}\mathsf{K}$.

The automaton $\Delta'(\mathsf{E})$ recognizes the language denoted by E (*cf.* [6, Th. 9]).[2] More precisely, it holds

Property 1.5. *For every* K *in* $\mathrm{BD}(\mathsf{E})$*, the future of* K *in* $\Delta'(\mathsf{E})$ *is equal to* $|\mathsf{K}|$.

The corrected version for Theorem 3.5 of [7] is then obtained by replacing *derived term* by *broken derived term*.

Theorem 1.6. *Let* $\mathcal{A}$ *be a co-deterministic automaton and* $\mathsf{E} = \Phi(\mathcal{A})$ *a rational expression computed from* $\mathcal{A}$ *by the state elimination method. Then, the* broken derived term *automaton* $\Delta'(\mathsf{E})$ *of* E *is co-deterministic.*

Figure 2 shows $\Delta'(\mathsf{E}_1)$, a co-deterministic automaton as expected, and whose minimal co-quotient is $\mathcal{A}_1$, as desired.

[2] We use that notation $\Delta'()$ here, in this corrigendum, in order to have a different notation from the one used in [7]; it might be the case that in further publications, we shall use A() which is more natural.

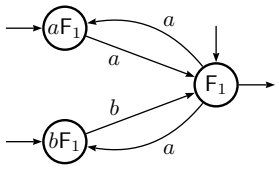


FIGURE 2. $\Delta'(\mathsf{E}_1)$

In [7], we have given a number of examples where the mapping $\Upsilon \circ \Delta'$ applied to an expression that is computed from a co-minimal automaton $\mathcal{A}$ yields $\mathcal{A}$ itself. But the only case where we have been able to prove the property is when $\mathcal{A}$ is co-deterministic. Moreover, this result gives the possibility of dealing with the general case by means of tagging. This is the reason why we consider Theorem 1.6 as the key statement of [7].

The introduction of the breaking operation is necessary to deal with automata with several initial states. Nevertheles, the core of the proof of Theorem 1.6 is Proposition 3.5 which establishes that the derived term automaton of an expression computed from a *normalized* co-deterministic automaton is co-deterministic, that is, a property of Antimirov derivation. A remarkable feature of this result is that its proof goes by induction on the number of states of the normalized automaton we start from. The main ingredients of this proof are first Lemma 2.1 that describes how the (Antimirov) derivation translates under a *continuous rational substitution* and then on Proposition 2.8 which states that every (broken) derived term is contained in the *future* of some state of the automaton on which the expression has been computed. The transfer from the hypothesis of Proposition 3.5 (normalized co-deterministic automata) to the one of Theorem 1.6 (co-deterministic automata) is made possible by Lemma 2.7.

2. Preliminary to the proof

We first establish two properties of derivation that will be used in the sequel and whose scope is wider than the statement alone. For the ease of the proof, we also rather work with *normalized* automata. We then show, in two steps, that we can make this assumption without loss of generality.

2.1. A substitution lemma

We call *rational substitution* a map $\varphi\colon B \to \mathsf{RatE}\, A^*$ which is extended to $\varphi\colon \mathsf{RatE}\, B^* \to \mathsf{RatE}\, A^*$ by replacing every atom of a rational expression over B^* by its image under φ. Substitutions are consistent with rational operations in the sense that for every $\mathsf{E}, \mathsf{F} \in \mathsf{RatE}\, B^*$ it holds:

$$\varphi(\mathsf{E}+\mathsf{F}) = \varphi(\mathsf{E}) + \varphi(\mathsf{F})\,, \quad \varphi(\mathsf{E}\cdot\mathsf{F}) = \varphi(\mathsf{E})\cdot\varphi(\mathsf{F}) \quad \text{and} \quad (\varphi(\mathsf{E}))^* = \varphi\left((\mathsf{E})^*\right)\,.$$

Naturally, φ also induces a substitution from B^* into A^* and it holds $|\varphi(\mathsf{E})| = \varphi(|\mathsf{E}|)$. A substitution φ is *continuous* if $\mathsf{c}(\varphi(b))$ is null for every b in B. If φ is *continuous*, it holds:

$$\forall \mathsf{E} \in \mathsf{RatE}\, B^* \qquad \mathsf{c}(\varphi(\mathsf{E})) = \mathsf{c}(\mathsf{E}) \ . \tag{6}$$

Lemma 2.1. *Let* $\varphi\colon \mathsf{RatE}\, B^* \to \mathsf{RatE}\, A^*$ *be a continuous rational substitution. Then*

$$\forall a \in A \qquad \frac{\partial}{\partial a}\varphi(\mathsf{E}) = \bigcup_{b\in B}\left[\frac{\partial}{\partial a}\varphi(b)\right]\varphi\left(\frac{\partial}{\partial b}\mathsf{E}\right) \ . \tag{7}$$

Proof. By induction on the depth of E. If $\mathsf{E} = \mathsf{0}$ or $\mathsf{1}$, both sides of (7) are zero; if $\mathsf{E} = b$, the right handside of (7) reduces to $\dfrac{\partial}{\partial a}\varphi(b)$ which is the left handside, and the base of the induction is established. The following three sequences of equalities give the three possible induction steps.

$$\begin{aligned}
\frac{\partial}{\partial a}\varphi(\mathsf{F}+\mathsf{G}) &= \frac{\partial}{\partial a}[\varphi(\mathsf{F})+\varphi(\mathsf{G})] = \frac{\partial}{\partial a}\varphi(\mathsf{F}) \cup \frac{\partial}{\partial a}\varphi(\mathsf{G}) \\
&= \bigcup_{b\in B}\left[\frac{\partial}{\partial a}\varphi(b)\right]\varphi\left(\frac{\partial}{\partial b}\mathsf{F}\right) \cup \bigcup_{b\in B}\left[\frac{\partial}{\partial a}\varphi(b)\right]\varphi\left(\frac{\partial}{\partial b}\mathsf{G}\right) \\
&= \bigcup_{b\in B}\left[\frac{\partial}{\partial a}\varphi(b)\right]\left[\varphi\left(\frac{\partial}{\partial b}\mathsf{F}\right) \cup \varphi\left(\frac{\partial}{\partial b}\mathsf{G}\right)\right] \\
&= \bigcup_{b\in B}\left[\frac{\partial}{\partial a}\varphi(b)\right]\varphi\left(\frac{\partial}{\partial b}(\mathsf{F}+\mathsf{G})\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial a}\varphi(\mathsf{F}\cdot\mathsf{G}) &= \frac{\partial}{\partial a}[\varphi(\mathsf{F})\cdot\varphi(\mathsf{G})] = \left[\frac{\partial}{\partial a}\varphi(\mathsf{F})\right]\varphi(\mathsf{G}) \cup \mathsf{c}(\varphi(\mathsf{F}))\ \frac{\partial}{\partial a}\varphi(\mathsf{G}) \\
&= \left[\bigcup_{b\in B}\left[\frac{\partial}{\partial a}\varphi(b)\right]\varphi\left(\frac{\partial}{\partial b}\mathsf{F}\right)\right]\cdot\varphi(\mathsf{G}) \cup \mathsf{c}(\mathsf{F}) \bigcup_{b\in B}\left[\frac{\partial}{\partial a}\varphi(b)\right]\varphi\left(\frac{\partial}{\partial b}\mathsf{G}\right) \\
&= \bigcup_{b\in B}\left[\frac{\partial}{\partial a}\varphi(b)\right]\varphi\left(\frac{\partial}{\partial b}(\mathsf{F}\cdot\mathsf{G})\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial a}\varphi(\mathsf{F}^*) &= \frac{\partial}{\partial a}(\varphi(\mathsf{F}))^* = \left[\frac{\partial}{\partial a}\varphi(\mathsf{F})\right]\cdot(\varphi(\mathsf{F}))^* \\
&= \left[\bigcup_{b\in B}\left[\frac{\partial}{\partial a}\varphi(b)\right]\varphi\left(\frac{\partial}{\partial b}\mathsf{F}\right)\right]\cdot(\varphi(\mathsf{F}))^* \\
&= \bigcup_{b\in B}\left[\frac{\partial}{\partial a}\varphi(b)\right]\left[\varphi\left(\frac{\partial}{\partial b}\mathsf{F}\right)\cdot(\varphi(\mathsf{F}))^*\right] = \bigcup_{b\in B}\left[\frac{\partial}{\partial a}\varphi(b)\right]\varphi\left(\frac{\partial}{\partial b}\mathsf{F}^*\right)
\end{aligned}$$

□

Note that Lemma 2.1 refers to the *derivation* and not to the breaking derivation; it turns out that we shall make use of the two notions in parallel.

2.2. Normalized automata

If $\mathcal{A}$ is an automaton over A, we write $|\mathcal{A}|$ for the language of A^* denoted by $\mathcal{A}$.

Definition 2.2. An automaton is *standard* if it has exactly one initial state with no incoming transition.

To any automaton $\mathcal{A} = \langle\, Q, A, E, I, T \,\rangle$ we associate the standard automaton $\mathcal{A}_\$ = \langle\, Q \cup \{i\}, A \cup \{\$\}, F, \{i\}, T \,\rangle$, with

$$F = E \cup \{(i, \$, p) \mid p \in I\}$$

and where $\$$ is not in A and i does not belong to Q. Note that $\mathcal{A}$ is co-deterministic if, and only if, so is $\mathcal{A}_\$$. Clearly $|\mathcal{A}_\$| = \$|\mathcal{A}|$. This equality indeed generalizes to the broken derived terms in the following way.

Lemma 2.3. *Let $\mathcal{A}$ be an automaton over A, $\mathcal{A}_\$$ the standard automaton defined as above, $\mathsf{E} = \Phi_\omega(\mathcal{A})$ and $\mathsf{E}_\$ = \Phi_\omega(\mathcal{A}_\$)$ the expressions obtained by the state elimination method (with respect to the same order ω on the states of $\mathcal{A}$). It then holds:*

$$\mathrm{d}(\mathsf{E}) = \frac{\partial_\mathsf{b}}{\partial \$}\mathsf{E}_\$ \qquad \textit{and} \qquad \mathrm{BD}(\mathsf{E}) = \mathrm{TBD}(\mathsf{E}_\$) \ .$$

Definition 2.4. An automaton is *co-standard* if it has exactly one final with no outgoing transition, that is, if its transposition is standard.

To any automaton $\mathcal{A} = \langle\, Q, A, E, I, T \,\rangle$ we associate the co-standard automaton $\mathcal{A}_\pounds = \langle\, Q \cup \{t\}, A \cup \{\pounds\}, F, I, \{t\} \,\rangle$, with

$$F = E \cup \{(p, \pounds, t) \mid p \in T\}$$

and where $\pounds$ is not in A and t does not belong to Q. Note that $\mathcal{A}$ is co-deterministic if, and only if, so is $\mathcal{A}_\pounds$. Clearly $|\mathcal{A}_\pounds| = |\mathcal{A}|\,\pounds$. This equality indeed generalizes to the broken derived terms in the following way.

Lemma 2.5. *Let $\mathcal{A}$ be an automaton over A, $\mathcal{A}_\pounds$ the co-standard automaton defined as above, $\mathsf{E} = \Phi_\omega(\mathcal{A})$ and $\mathsf{E}_\pounds = \Phi_\omega(\mathcal{A}_\pounds)$. The projection $\pi\colon (A \cup \{\pounds\})^* \longrightarrow A^*$ induces a bijection between $\mathrm{BD}(\mathsf{E}_\pounds) \setminus \{1\}$ and $\mathrm{BD}(\mathsf{E})$ and in particular between $\mathrm{d}(\mathsf{E}_\pounds)$ and $\mathrm{d}(\mathsf{E})$.*

Proof. Let $\mathcal{F}_\pounds$ be the family of expressions in $\mathsf{RatE}\,(A \cup \pounds)$ that denote rational subsets of A^* followed by $\pounds$:

$$\mathcal{F}_\pounds = \{\mathsf{E} \in \mathsf{RatE}\,(A \cup \pounds) \mid |E| = L\,\pounds, L \in \mathrm{Rat}\,A^*\} \ .$$

Since any word in the language denoted by an element of $\mathcal{F}_\pounds$ does not contain the symbol $\pounds$ but at the end, any element of $\mathcal{F}_\pounds$ is of one of the following forms: $\pounds$, $\mathsf{E}_\pounds + \mathsf{F}_\pounds$, or $\mathsf{G}\mathsf{F}_\pounds$, where $\mathsf{E}_\pounds$ and $\mathsf{F}_\pounds$ are in $\mathcal{F}_\pounds$ and G is in $\mathsf{RatE}\,A$.

As the algorithm Φ acts symbolically on the labels of the transitions, $\pi(\mathsf{E}_{\pounds}) = \mathsf{E}$.

By induction on the depth of the expression, we first show that $\pi(\mathrm{d}(\mathsf{E}_{\pounds})) = \mathrm{d}(\pi(\mathsf{E}_{\pounds})) = \mathrm{d}(\mathsf{E})$. The interesting case is when $\mathsf{E}_{\pounds} = \mathsf{G}\mathsf{F}_{\pounds}$:

$$\pi(\mathrm{d}(\mathsf{G}\mathsf{F}_{\pounds})) = \pi(\mathrm{d}(\mathsf{G})\ \mathsf{F}_{\pounds}) \cup \delta_{\mathrm{d}(\mathsf{G})}\, \pi(\mathrm{d}(\mathsf{F}_{\pounds})) = \mathrm{d}(\mathsf{G})\ \mathsf{F} \cup \delta_{\mathrm{d}(\mathsf{G})}\ \mathrm{d}(\mathsf{F})) = \mathrm{d}(\mathsf{G}\mathsf{F}) \quad .$$

In the same way, for any $\mathsf{H}_{\pounds}$ in $\mathcal{F}_{\pounds}$, and for any letter a in A, it holds:

$$\pi(\frac{\partial_{\mathsf{b}}}{\partial a}\,\mathsf{H}_{\pounds}) = \frac{\partial_{\mathsf{b}}}{\partial a}\,\pi(\mathsf{H}_{\pounds}) \quad . \tag{8}$$

The statement is then shown by induction on the length of the derivation. Every (broken) derived term of $\mathsf{E}_{\pounds}$ different from 1 is in $\mathcal{F}_{\pounds}$. Let $\mathsf{K}_{\pounds}$ be a derived term in $\mathrm{BD}(\mathsf{E}_{\pounds})$ and K its projection, in $\mathrm{BD}(\mathsf{E})$ by induction hypothesis. The derivation of $\mathsf{K}_{\pounds}$ with respect to $\pounds$ gives either 1, or the empty set. The derivation of $\mathsf{K}_{\pounds}$ with respect to any letter a in A is, by (8), in bijection with the derivation of K with respect to a. □

Putting the standardisation and co-standardisation together, we get the following definition and statement.

Definition 2.6. An automaton is *normalized* if and only if it has exactly one initial state and one final state such that there is no incoming transition on the initial state and no outgoing transition from the final state.

To any automaton $\mathcal{A} = \langle\, Q, A, E, I, T \,\rangle$ we associate the normalized automaton $\mathcal{A}_{\mathsf{n}} = \langle\, Q \cup \{i, t\}, A \cup \{\$, \pounds\}, F, \{i\}, \{t\} \,\rangle$, with

$$F = E \cup \{(i, \$, p) \mid p \in I\} \cup \{(p, \pounds, t) \mid p \in T\}$$

and where $\$$ and $\pounds$ are not in A and i and t do not belong to Q. Note that $\mathcal{A}$ is co-deterministic if, and only if, so is $\mathcal{A}_{\mathsf{n}}$. Clearly $|\mathcal{A}_{\mathsf{n}}| = \$|\mathcal{A}|\,\pounds$. This equality indeed generalizes to the broken derived terms in the following way.

Lemma 2.7. *Let $\mathcal{A}$ be an automaton over A, $\mathcal{A}_{\mathsf{n}}$ the normalized automaton defined as above, $\mathsf{E} = \Phi_\omega(\mathcal{A})$ and $\mathsf{E}_{\mathsf{n}} = \Phi_\omega(\mathcal{A}_{\mathsf{n}})$. The projection $\pi\colon (A \cup \{\pounds\})^* \longrightarrow A^*$ induces a bijection between $\dfrac{\partial_{\mathsf{b}}}{\partial \$}\,\mathsf{E}_{\mathsf{n}}$ and $\mathrm{d}(\mathsf{E})$ on one hand and between $\mathrm{TBD}(\mathsf{E}_{\mathsf{n}}) \setminus \{1\}$ and $\mathrm{BD}(\mathsf{E})$ on the other hand.* □

2.3. Future of states and derived terms

Let $\mathcal{A} = \langle\, Q, A, E, I, T \,\rangle$ be an automaton over A^*. For each state q of $\mathcal{A}$, the *past of q (in $\mathcal{A}$)* is the set of labels of computation which go from an initial state of $\mathcal{A}$ to q, and we write it $\mathsf{Past}_{\mathcal{A}}(q)$; the *future of q (in $\mathcal{A}$)* is the set of labels of computations that go from q to a final state of $\mathcal{A}$ and we write it $\mathsf{Fut}_{\mathcal{A}}(q)$:

$$\mathsf{Past}_{\mathcal{A}}(q) = \{w \in A^* \mid \exists i \in I \quad i \xrightarrow[\mathcal{A}]{w} q\}, \quad \mathsf{Fut}_{\mathcal{A}}(q) = \{w \in A^* \mid \exists t \in T \quad q \xrightarrow[\mathcal{A}]{w} t\}.$$

An automaton $\mathcal{A}$ is *deterministic* if, and only if, the pasts of states are pairwise disjoint, and dually, $\mathcal{A}$ is *co-deterministic* if, and only if, the futures of states are pairwise disjoint.

Proposition 2.8. *Let $\mathcal{A}$ be a* standard *automaton with set of states Q and $\mathsf{E} = \Phi(\mathcal{A})$. For every derived term K of E there exists a state q in Q such that $|\mathsf{K}|$ is included in the future of q.*

Proof. We first transform $\mathcal{A} = \langle Q, A, E, I, T \rangle$ into an automaton $\widehat{\mathcal{A}} = \left\langle Q, A\times Q, \widehat{E}, I, T \right\rangle$ by relabelling every transition (p, a, q) in E as (p, a_q, q). The projection π maps $A\times Q$ onto A and is extended to a map from $\mathsf{RatE}\,(A\times Q)$ onto $\mathsf{RatE}\,A$.

As π is a continuous (rational) substitution, with the property that for any a, b in A, (and any q in Q), $\dfrac{\partial}{\partial a}\,\pi(b_q)$ is equal to 1 or 0 according to whether a is equal to b or not, Lemma 2.1 implies that for any a in A and for any expression F in $\mathsf{RatE}\,(A\times Q)$ it holds

$$\frac{\partial}{\partial a}\,\pi(\mathsf{F}) = \bigcup_{q\in Q}\left[\pi\left(\frac{\partial}{\partial a_q}\,\mathsf{F}\right)\right] \quad . \tag{9}$$

Let $\widehat{\mathsf{E}} = \Phi(\widehat{\mathcal{A}})$. As the algorithm Φ acts symbolically on the labels of the transitions to build the expression, it holds: $\pi(\widehat{\mathsf{E}}) = \mathsf{E}$. Then, by iteration of (9), for any K in $\mathrm{D}(E)$ there exists K' in $\mathrm{D}(\widehat{E})$ such that $\mathsf{K} = \pi(\mathsf{K}')$. Thus, there exists w in $(A\times Q)^*$ such that K' is in $\dfrac{\partial}{\partial w}\,\widehat{E}$ and then

$$|\mathsf{K}'| \subseteq |\frac{\partial}{\partial w}\,\widehat{E}| = w^{-1}|\widehat{\mathcal{A}}| \quad .$$

If $\mathsf{K}' = \widehat{E}$, then $\mathsf{K} = \mathsf{E}$ and $|\mathsf{K}'|$ is (in) the future of the (unique) initial state. Otherwise, w is a non empty word; let a_q is the last letter of w, $w^{-1}|\widehat{\mathcal{A}}|$ is contained in $\mathsf{Fut}_{\widehat{\mathcal{A}}}(q)$. The projection by π gives

$$\mathsf{Fut}_{\mathcal{A}}(q) = \pi\left(\mathsf{Fut}_{\widehat{\mathcal{A}}}(q)\right) \qquad \text{and thus} \qquad |\mathsf{K}| \subseteq \mathsf{Fut}_{\mathcal{A}}(q) \quad .$$

□

The same property holds then for broken derived terms.

Corollary 2.9. *Let $\mathcal{A}$ be an automaton with set of states Q and $\mathsf{E} = \Phi_\omega(\mathcal{A})$. For every* broken *derived term K of E there exists a state q in Q such that $|\mathsf{K}|$ is included in the future of q.*

Proof. Let $\mathcal{A}_\$$ be the standard automaton associated with $\mathcal{A}$ as in Section 2.2 and let $\mathsf{E}_\$ = \Phi_\omega(\mathcal{A}_\$)$. For every $\mathsf{K}_\$$ in $\mathrm{BD}(\mathsf{E}_\$)$ there exists $\mathsf{H}_\$$ in $\mathrm{D}(\mathsf{E}_\$)$ such that $\mathsf{K}_\$$ is in $\mathrm{d}(\mathsf{H}_\$)$. Moreover, there exists, by Proposition 2.8, a state q in Q such that

$|\mathsf{H}_\$| \subseteq \mathsf{Fut}_{\mathcal{A}_\$}(q)$ hence $|\mathsf{K}_\$| \subseteq \mathsf{Fut}_{\mathcal{A}_\$}(q)$. By Lemma 2.3, $\mathrm{TBD}(\mathsf{E}_\$) = \mathrm{BD}(\mathsf{E})$, therefore $|\mathsf{K}| \subseteq \mathsf{Fut}_{\mathcal{A}}(q)$. □

As the futures of states are pairwise disjoint in a co-deterministic automaton, we then can state the following.

Corollary 2.10. *Let $\mathcal{A}$ be a* co-deterministic *normalized automaton with set of states Q and $\mathsf{E} = \Phi(\mathcal{A})$. For every K of $\mathrm{D}(\mathsf{E})$, there exists a* unique *q in Q such that $|\mathsf{K}| \subseteq \mathsf{Fut}_{\mathcal{A}}(q)$; for every L of $\mathrm{BD}(\mathsf{E})$, there exists a* unique *q in Q such that $|\mathsf{L}| \subseteq \mathsf{Fut}_{\mathcal{A}}(q)$.* □

3. Proof of Theorem 1.6

The proof goes by induction on the number of states of the automaton $\mathcal{A}$ as sketched in Figure 3.

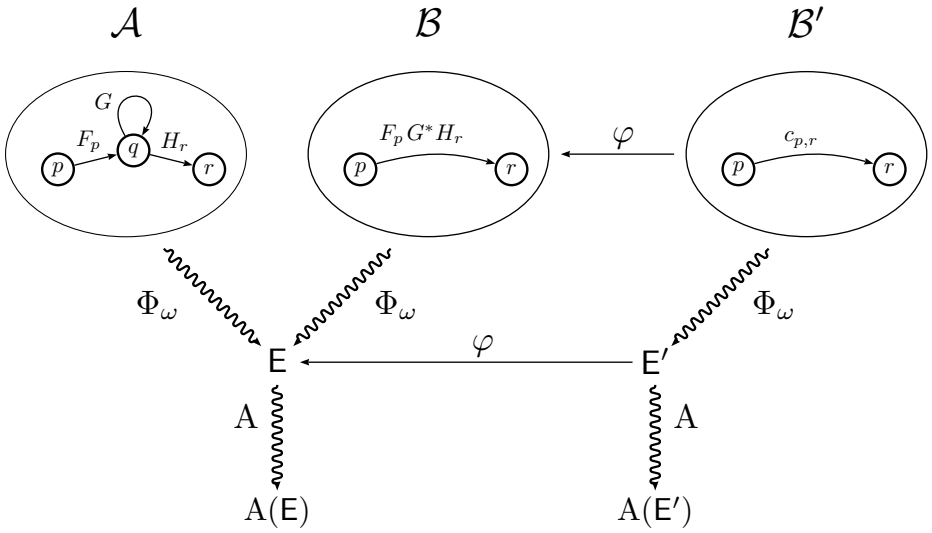


Figure 3. Diagram for an induction step in the proof of Theorem 1.6

Let $\mathsf{E} = \Phi_\omega(\mathcal{A})$ the rational expression computed by the state elimination method following the order ω on the states of $\mathcal{A}$ and let q be the smallest state with respect to ω. For every predecessor p of q, let F_p be the set of letters a such that (p, a, q) is a transition of $\mathcal{A}$ and F_p the expression 'sum of letters in F_p'. Let R be the set of successors of q in $\mathcal{A}$; for every r in R, let H_r be the set of letters a such that (q, a, r) is a transition of $\mathcal{A}$ and H_r the expression 'sum of letters in H_r'. Let G be the set of letters a such that (q, a, q) is a transition of $\mathcal{A}$ and G the expression 'sum of letters in G' (G may be empty and G null).

Let $\mathcal{B}$ be the *generalized* automaton obtained from $\mathcal{A}$ by elimination of q — by the state elimination method. For every p and every r as above, a transition

from p to r labelled by $\mathsf{F}_p\mathsf{G}^*\mathsf{H}_r$ (*resp.* $\mathsf{F}_p\mathsf{H}_r$ if G is null) is thus created.[3] Let $\mathcal{B}'$ be the (normalized) automaton obtained from $\mathcal{B}$ by replacing every label of the form $\mathsf{F}_p\mathsf{G}^*\mathsf{H}_r$ (or $\mathsf{F}_p\mathsf{H}_r$) by a fresh letter $c_{p,r}$. Let C be the alphabet of these fresh letters and $A' = A \cup C$. Let φ be the (continuous) substitution from $\mathsf{RatE}\, A'^*$ to $\mathsf{RatE}\, A^*$ which maps every letter of A onto itself and every letter $c_{p,r}$ onto $\mathsf{F}_p\mathsf{G}^*\mathsf{H}_r$ (or $\mathsf{F}_p\mathsf{H}_r$ if G is null).

Since the construction of $\mathcal{B}$ is the first step of $\Phi_\omega()$, then $\mathsf{E} = \Phi_\omega(\mathcal{B})$ and let $\mathsf{E}' = \Phi_\omega(\mathcal{B}')$. As the algorithm Φ_ω acts symbolically on the labels of the transitions, $\varphi(\mathsf{E}') = \mathsf{E}$.

If $\mathcal{A}$ is co-deterministic, so is $\mathcal{B}'$ and $\Delta'(\mathsf{E}')$ is co-deterministic by induction hypothesis. The proof amounts to transfer the properties from $\Delta'(\mathsf{E}')$ to $\Delta'(\mathsf{E})$ via the substitution φ.

3.1. Preparation for the induction

From now on — and until after Corollary 3.6 — $\mathcal{A}$, and thus $\mathcal{B}$ and $\mathcal{B}'$, are *normalized* automaton. The state elimination method doest not eliminate neither the initial nor the final state and thus every state is treated in the same way. With the above notation, it first holds:

Property 3.1. *For every state s, $s \neq q$, $\mathsf{Fut}_{\mathcal{A}}(s) = \varphi\left(\mathsf{Fut}_{\mathcal{B}'}(s)\right)$.* □

The remainder of this subsection is devoted to the description of the image of the broken derived terms under φ. From Lemma 2.1, for every K' in $\mathsf{RatE}\, A'^*$, and every letter a in A, it holds

$$\frac{\partial}{\partial a}\varphi(\mathsf{K}') = \varphi\left(\frac{\partial}{\partial a}\mathsf{K}'\right) \cup \bigcup_{c\in C}\left[\frac{\partial}{\partial a}\varphi(c)\right]\varphi\left(\frac{\partial}{\partial c}\mathsf{K}'\right) \quad .$$

For every c in C, there exists p and r such that $\varphi(c) = \mathsf{F}_p\mathsf{G}^*\mathsf{H}_r$ (*resp.* $\mathsf{F}_p\mathsf{H}_r$). Hence, either a does not belong to any F_p and $\dfrac{\partial}{\partial a}\varphi(\mathsf{K}') = \varphi\left(\dfrac{\partial}{\partial a}\mathsf{K}'\right)$ or it belongs to F_p for a unique p and (recall that R is the set of successors of q in $\mathcal{A}$):

$$\frac{\partial}{\partial a}\varphi(\mathsf{K}') = \begin{cases} \varphi\left(\dfrac{\partial}{\partial a}\mathsf{K}'\right) \cup \bigcup\limits_{r\in R} \mathsf{G}^*\mathsf{H}_r\,\varphi\left(\dfrac{\partial}{\partial c_{p,r}}\mathsf{K}'\right) & \text{if } \mathsf{G} \text{ is not empty,} \\ \varphi\left(\dfrac{\partial}{\partial a}\mathsf{K}'\right) \cup \bigcup\limits_{r\in R} \mathsf{H}_r\,\varphi\left(\dfrac{\partial}{\partial c_{p,r}}\mathsf{K}'\right) & \text{otherwise.} \end{cases} \tag{10}$$

As G and H_r are sums of letters, for every c in C the derivation of $\mathsf{G}^*\mathsf{H}_r\,\varphi\left(\dfrac{\partial}{\partial c}\mathsf{K}'\right)$ (*resp.* $\mathsf{H}_r\,\varphi\left(\dfrac{\partial}{\partial c}\mathsf{K}'\right)$) with respect to any letter gives $\emptyset$, $\mathsf{G}^*\mathsf{H}_r\,\varphi\left(\dfrac{\partial}{\partial c}\mathsf{K}'\right)$ or

[3] The automaton $\mathcal{B}$ is labelled by rational expressions, which is the reason why it is called 'generalized'.

$\varphi\left(\dfrac{\partial}{\partial c}\,\mathsf{K}'\right)$. Hence, by a straightforward induction, it holds :

$$\mathrm{D}(\mathsf{E}) = \varphi\left(\mathrm{D}(\mathsf{E}')\right) \cup \bigcup_{c_{p,r}\in C}\left[\bigcup_{a\in F_p}\frac{\partial}{\partial a}\,\varphi(c_{p,r})\right]\varphi(\mathrm{D}_{c_{p,r}}(\mathsf{E}'))\ . \tag{11}$$

By applying d() to (10), it comes $\dfrac{\partial_{\mathsf{b}}}{\partial a}\,\varphi(\mathsf{K}') = \varphi\left(\dfrac{\partial_{\mathsf{b}}}{\partial a}\,\mathsf{K}'\right)$ in the former case, and in the latter:

$$\frac{\partial_{\mathsf{b}}}{\partial a}\,\varphi(\mathsf{K}') = \begin{cases} \varphi\left(\dfrac{\partial_{\mathsf{b}}}{\partial a}\,\mathsf{K}'\right) \cup \displaystyle\bigcup_{r\in R}\mathsf{G}^*\mathsf{H}_r\,\varphi\left(\frac{\partial}{\partial c_{p,r}}\,\mathsf{K}'\right) & \text{if } \mathsf{G} \text{ is not empty,} \\ \varphi\left(\dfrac{\partial_{\mathsf{b}}}{\partial a}\,\mathsf{K}'\right) \cup \displaystyle\bigcup_{r\in R}\bigcup_{b\in H_r} b\,\varphi\left(\frac{\partial}{\partial c_{p,r}}\,\mathsf{K}'\right) & \text{otherwise.} \end{cases} \tag{12}$$

As above, and for every c in C the *breaking* derivation of $\mathsf{G}^*\mathsf{H}_r\,\varphi\left(\dfrac{\partial}{\partial c}\,\mathsf{K}'\right)$ (*resp.* $b\,\varphi\left(\dfrac{\partial}{\partial c}\,\mathsf{K}'\right)$) with respect to any letter gives $\emptyset$, $\mathsf{G}^*\mathsf{H}_r\,\varphi\left(\frac{\partial}{\partial c}\,\mathsf{K}'\right)$ or $\varphi\left(\dfrac{\partial_{\mathsf{b}}}{\partial c}\,\mathsf{K}'\right)$. Hence, by a straightforward induction, it holds :

$$\mathrm{BD}(\mathsf{E}) = \varphi\left(\mathrm{BD}(\mathsf{E}')\right) \cup \bigcup_{c_{p,r}\in C}\left[\bigcup_{a\in F_p}\frac{\partial_{\mathsf{b}}}{\partial a}\,\varphi(c_{p,r})\right]\varphi(\mathrm{D}_{c_{p,r}}(\mathsf{E}'))\ . \tag{13}$$

3.2. Proof of Theorem 1.6

The comparison between (11) and (13) is quite instructive. The presence in the right handside of both equations of a *non breaking derivation* makes it understandable that an induction proof will run more smoothly on the derived term automata. This is done with Proposition 3.5. On the other hand, when it comes to compare the derived terms of an expression obtained from an automaton $\mathcal{A}$ and those obtained from its normalization $\mathcal{A}_{\$}$, the *broken derived terms* are to be used (Lemma 2.7). Corollary 3.6 will fill the gap and Theorem 1.6 will follow.

We keep the notation of the preceeding subsection and we suppose from now on that $\mathcal{A}$ is (normalized) *co-deterministic*. This implies that $\mathcal{B}'$ is co-deterministic and moreover that for every p, G and F_p are disjoint; likewise, if p and p' are two distinct predecessors of q, F_p and $F_{p'}$ are disjoint. For every successor r of q, let I_r be the set of labels of incoming transitions of r that does not come from q; then I_r and H_r are disjoint. For every state s of $\mathcal{B}'$ (therefore different from q), let J_s be the set of labels of incoming transitions of s. It then holds:

Property 3.2. *For every state s in $\mathcal{B}'$, $\varphi(J_s)$ is a suffix code in A^* and the images of the elements of J_s by φ are pairwise disjoint.*

Lemma 3.3. *For every state s in $\mathcal{A}$, $s \neq q$, and for every u in $\mathsf{Fut}_{\mathcal{A}}(s)$ there exists a unique v in $\mathsf{Fut}_{\mathcal{B}'}(s)$ such that $u \in \varphi(v)$.*

Proof. Suppose by way of contradiction that there exist v and v' in $\mathsf{Fut}_{\mathcal{B}'}(s)$ such that $u \in \varphi(v) \cap \varphi(v')$. We write $v = x_n \dots x_2 x_1$ and $v' = x'_m \dots x'_2 x'_1$. There exist $w_i \in \varphi(x_i)$ and $w'_i \in \varphi(x'_i)$ such that $u = w_n \dots w_2 w_1 = w'_m \dots w'_2 w'_1$.

Let j be the smallest index such that either $w_j \neq w'_j$ or $x_j \neq x'_j$. Let r be the unique state of $\mathcal{B}'$ such that $x_{j-1} \dots x_1 \in \mathsf{Fut}_{\mathcal{B}'}(r)$. By Property 3.2, $\varphi(J_r)$ is a suffix code and then $w_j = w'_j$, and $\varphi(x_j) \cap \varphi(x'_j) \neq \emptyset$ implies $x_j = x'_j$. □

Let K be a derived term in $\mathrm{D}(\mathsf{E})$. As $\mathcal{A}$ is co-deterministic, and by Proposition 2.8 either there exists a *unique* state $s \neq q$ in $\mathcal{A}$ such that $|\mathsf{K}| \subseteq \mathsf{Fut}_{\mathcal{A}}(s)$ — **Case 1** — or $|\mathsf{K}| \subseteq \mathsf{Fut}_{\mathcal{A}}(q)$ — **Case 2**.

Lemma 3.4. *Let K in $\mathrm{D}(\mathsf{E})$. There exists K' in $\mathrm{D}(\mathsf{E}')$ such that either:*
Case 1 *there exists $s \neq q$ such that $|\mathsf{K}'| \subseteq \mathsf{Fut}_{\mathcal{B}'}(s)$ and $\mathsf{K} = \varphi(\mathsf{K}')$, or*
Case 2 *there exists a successor r of q such that $|\mathsf{K}'| \subseteq \mathsf{Fut}_{\mathcal{B}'}(r)$ and $\mathsf{K} = \mathsf{G}^*\mathsf{H}_r\varphi(\mathsf{K}')$ or $\mathsf{K} = \mathsf{H}_r\varphi(\mathsf{K}')$ (if G is null).*

Proof. From (11), K is either (**Case 1**) in $\varphi(\mathrm{D}(\mathsf{E}'))$ or (**Case 2**) in

$$\bigcup_{c_{p,r} \in C} \left[\bigcup_{a \in F_p} \frac{\partial}{\partial a} \varphi(c_{p,r}) \right] \varphi(\mathrm{D}_{c_{p,r}}(\mathsf{E}')) \ .$$

In **Case 1**, there exists K' such that $\mathsf{K} = \varphi(\mathsf{K}')$, by Proposition 2.8 there exists s in $\mathcal{B}'$ such that $|\mathsf{K}'| \subseteq \mathsf{Fut}_{\mathcal{B}'}(s)$ and by Property 3.1 $|\mathsf{K}| \subseteq \mathsf{Fut}_{\mathcal{A}}(s)$.

In **Case 2**, there exist $c_{p,r} \in C$, $a \in A$ and $\mathsf{K}' \in \mathrm{D}_{c_{p,r}}(\mathsf{E}')$ such that

$$\mathsf{K} = \left[\frac{\partial}{\partial a} \varphi(c_{p,r}) \right] \varphi(\mathsf{K}') = \mathsf{G}^*\mathsf{H}_r\varphi(\mathsf{K}') \ .$$

As $\mathsf{K}' \in \mathrm{D}_{c_{p,r}}(\mathsf{E}')$, $|\mathsf{K}'| \subseteq \mathsf{Fut}_{\mathcal{B}'}(r)$. By Property 3.1 $\varphi(|\mathsf{K}'|) \subseteq \mathsf{Fut}_{\mathcal{A}}(r)$ and thus, by definition of G and H_r, $|\mathsf{K}| \subseteq \mathsf{Fut}_{\mathcal{A}}(q)$. □

Proposition 3.5. *Let $\mathcal{A}$ be a normalized co-deterministic automaton, and $\mathsf{E} = \Phi(\mathcal{A})$. Then, the derived term automaton $\Delta(\mathsf{E})$ of E is co-deterministic.*

Proof. A normalized automaton has at least two states. If $\mathcal{A}$ has only two states, $\Phi(\mathcal{A})$ is reduced to a sum of letters and $\Delta(\Phi(\mathcal{A}))$ is clearly co-deterministic.

By induction, $\Delta(\mathsf{E}')$ is co-deterministic, which means, by Property 1.5, that the interpretations of the derived terms of E' are pairwise disjoint. We prove now that the interpretations of elements of $\mathrm{D}(\mathsf{E})$ are disjoint, which implies that $\Delta(\mathsf{E})$ is co-deterministic.

Let K_1 and K_2 be two distinct derived terms of $\mathrm{D}(\mathsf{E})$ and assume that there exists u in $|\mathsf{K}_1| \cap |\mathsf{K}_2|$. By (11) and Lemma 3.4, there exist K'_1 and K'_2 in $\mathrm{D}(\mathsf{E}')$ such that one of the following three cases holds:
Case 1 $\mathsf{K}_1 = \varphi(\mathsf{K}'_1)$ and $\mathsf{K}_2 = \varphi(\mathsf{K}'_2)$.

Case 2.1 $\mathsf{K}_1 = \varphi(\mathsf{K}'_1)$ and $\mathsf{K}_2 = \mathsf{G}^*\mathsf{H}_r\varphi(\mathsf{K}'_2)$.
Case 2.2 $\mathsf{K}_1 = \mathsf{G}^*\mathsf{H}_{r_1}\varphi(\mathsf{K}'_1)$ and $\mathsf{K}_2 = \mathsf{G}^*\mathsf{H}_{r_2}\varphi(\mathsf{K}'_2)$.

Case 1 From Lemma 3.4 there exist states s_1 and s_2 such that $|\mathsf{K}'_1| \subseteq \mathsf{Fut}_{\mathcal{B}'}(s_1)$ and $|\mathsf{K}'_2| \subseteq \mathsf{Fut}_{\mathcal{B}'}(s_2)$, and $|\mathsf{K}_1| \subseteq \mathsf{Fut}_{\mathcal{A}}(s_1)$ and $|\mathsf{K}_2| \subseteq \mathsf{Fut}_{\mathcal{A}}(s_2)$. As $\mathcal{A}$ is co-deterministic, $s_1 = s_2 = s \neq q$ and by Lemma 3.3, there exists a unique v in $\mathsf{Fut}_{\mathcal{B}'}(s)$ such that $u \in \varphi(v)$. Hence $v \in |\mathsf{K}'_1| \cap |\mathsf{K}'_2|$. The induction hypothesis implies $\mathsf{K}'_1 = \mathsf{K}'_2$ and thus $\mathsf{K}_1 = \mathsf{K}_2$.

Case 2.1 From Lemma 3.4 there exist states $s_1 \neq q$ such that $|\mathsf{K}_1| \subseteq \mathsf{Fut}_{\mathcal{A}}(s_1)$ and $|\mathsf{K}_2| \subseteq \mathsf{Fut}_{\mathcal{A}}(q)$. As $\mathcal{A}$ is co-deterministic, $|\mathsf{K}_1| \cap |\mathsf{K}_2| = \emptyset$.

Case 2.2 From Lemma 3.4, there exist successors of q, r_1 and r_2 such that $|\mathsf{K}'_1| \subseteq \mathsf{Fut}_{\mathcal{B}'}(r_1)$ and $|\mathsf{K}'_2| \subseteq \mathsf{Fut}_{\mathcal{B}'}(r_2)$, and $|\mathsf{K}_1| \subseteq \mathsf{Fut}_{\mathcal{A}}(q)$ and $|\mathsf{K}_2| \subseteq \mathsf{Fut}_{\mathcal{A}}(q)$. Hence $u = g_1h_1w_1 = g_2h_2w_2$, with $g_1, g_2 \in G^*$, $h_1 \in H_{r_1}$, $h_2 \in H_{r_2}$, $w_1 \in \varphi(|\mathsf{K}'_1|)$ and $w_2 \in \varphi(|\mathsf{K}'_2|)$. As $G \cap H_{r_1} = G \cap H_{r_2} = \emptyset$, $g_1 = g2$, $h_1 = h2$, and thus $w_1 = w_2$. Since $\mathcal{A}$ is co-deterministic, there is a unique state s, different from q, such that w is in $\mathsf{Fut}_{\mathcal{A}}(s)$. Hence, by Case 1, $\varphi(\mathsf{K}'_1) = \varphi(\mathsf{K}'_2)$ and thus $\mathsf{K}_1 = \mathsf{K}_2$. □

Corollary 3.6. *Let $\mathcal{A}$ be a normalized co-deterministic automaton, and $\mathsf{E} = \Phi(\mathcal{A})$. Then, the broken derived term automaton $\Delta'(\mathsf{E})$ of E is co-deterministic.*

Proof. As above, the corollary trivially holds if $\mathcal{A}$ has only two states.

By induction, $\Delta'(\mathsf{E}')$ is co-deterministic, which means, by Property 1.5, that the interpretations of the broken derived terms of E' are pairwise disjoint. We prove now that the interpretations of elements of $\mathrm{BD}(\mathsf{E})$ are disjoint, which implies that $\Delta'(\mathsf{E})$ is co-deterministic.

By Equation (5), $\mathrm{BD}(\mathsf{E}) = \mathrm{d}(\mathrm{D}(\mathsf{E}))$ and, by Proposition 3.5, the interpretations of the derived terms of E are disjoint. If there exists a word u in the intersection of the interpretation of two broken derived terms K_1 and K_2, then, there exists a *derived term* L of E, such that both K_1 and K_2 are in $\mathrm{d}(\mathsf{L})$.

Case 1 There exists L' in $\mathrm{D}(\mathsf{E}')$ such that $\mathsf{L} = \varphi(\mathsf{L}')$. Then, there exists K'_1 and K'_2 in $\mathrm{d}(\mathsf{L}')$ such that $\mathsf{K}_1 = \varphi(\mathsf{K}'_1)$ and $\mathsf{K}_2 = \varphi(\mathsf{K}'_2)$. By Lemma 3.4 and Lemma 3.3, there exists a unique word v in $|\mathsf{L}'|$ such that $\varphi(v) = u$, hence, v is in $|\mathsf{K}'_1| \cap |\mathsf{K}'_2|$. By induction hypothesis, broken derived terms of E' have disjoint interpretations, thus $\mathsf{K}'_1 = \mathsf{K}'_2$ and $\mathsf{K}_1 = \mathsf{K}_2$.

Case 2 There exist L' in $\mathrm{D}(\mathsf{E}')$, G and H_r, such that $\mathsf{L} = \mathsf{G}^*\mathsf{H}_r\varphi(\mathsf{L}')$ or $\mathsf{L} = \mathsf{H}_r\varphi(\mathsf{L}')$ (if G is null).

In the first subcase, $\mathrm{d}(\mathsf{L}) = \{\mathsf{L}\}$, hence $\mathsf{K}_1 = \mathsf{K}_2 = \mathsf{L}$; in the second subcase, $\mathsf{K}_1 = a_1\varphi(\mathsf{L}')$ and $\mathsf{K}_2 = a_2\varphi(\mathsf{L}')$, a_1 and a_2 are both the first letter of u and are therefore equal, hence $\mathsf{K}_1 = \mathsf{K}_2$. □

Theorem 1.6 is now established. Indeed, either $\mathcal{A}$ is normalized, and nothing is to be added, or we apply Corollary 3.6 to $\mathcal{A}_\$$ and as the interpretations of any two broken derived terms of $\mathsf{E}_\$$ are disjoint so are the corresponding broken derived terms of E: the broken derived term automaton is co-deterministic. □

Conclusion

As we wanted to be complete, and not leave another flaw behind us, the proof of the theorem is long, much longer than we expected, and somewhat laboured. It goes by a repeated forth and back between the properties of the terms and of their interpretation. It is also based on an interplay between the derived terms and the broken ones, which let us think that the truth may be hidden somewhere in between...

In any case, this proof brought under light these broken terms that we defined in [6] in an exploratory way. They deserve to be further studied, a task that we have already begun [1].

Acknowledgements The authors are pleased to thanks Florent Terrones, from the Vaucanson Group, who pointed out an example which led to this correction and to Pierre-Yves Angrand who drew our attention on a problem with the breaking operation d() and gave the correct definition.

References

[1] P.-Y. Angrand, S. Lombardy and J. Sakarovitch, On the broken derived terms of a rational expression. *In preparation.*

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

[3] J. A. Brzozowski, Derivatives of regular expressions. *J. Assoc. Comput. Mach.* **11** (1964), 481–494.

[4] P. Caron and M. Flouret, Glushkov construction for series: the non commutative case, *Int. J. Comput. Math.* **80** (2003), 457–472.

[5] V. Glushkov, The abstract theory of automata. *Russian Mathematical Surveys* **16** (1961), 1–53.

[6] S. Lombardy and J. Sakarovitch, Derivatives of rational expressions with multiplicity, *Theoret. Computer Sci.* **332** (2005), 141–177. (Journal version of *Proc. MFCS 02*, LNCS 2420 (2002), 471–482.)

[7] S. Lombardy and J. Sakarovitch, How expressions can code for automata, *RAIRO – Theoret. Informatics and Applications* **39** (2005), 217–237. (Journal version of *Proc. of LATIN 2004* , LNCS 2976 (2004), 242–251.)

.