Multiplicités

Le but de ce chapitre est de généraliser les études et les résultats précédemment présentés au cas des expressions rationnelles à multiplicités dans un semi-anneau $\mathbb{K}$. La première chose qu'il faut faire, pour manipuler les expressions rationnelles à multiplicités, est de définir le cadre dans lequel on peut les considérer. En effet, l'étoile d'un scalaire de $\mathbb{K}$ n'étant pas nécessairement définie, il faut s'assurer tout d'abord que la notion d'étoile soit cohérente dans le semi-anneau $\mathbb{K}$. Ensuite il faut également s'assurer qu'étant donné une expression sur $\mathbb{K}$, les étoiles des scalaires soient définies.

Les termes dérivés d'une expression rationnelle à multiplicités ont été définis dans [31] afin de pouvoir construire un $\mathbb{K}$-automate qui soit un quotient de l'automate des positions de l'expression. Comme pour le cas booléen, les termes dérivés sont définis conjointement à une opération de dérivation. Cependant la relation entre termes dérivés et cette dérivation n'est pas aussi directe que dans le cas booléen. En effet la dérivation à multiplicités ne produit pas des ensembles d'expressions rationnelles mais des combinaisons linéaires – c'est-à-dire des ensembles dont chaque élément est pondéré par un élément de $\mathbb{K}$ – d'expression rationnelles. Par le jeu des coefficients qui peuvent s'annuler, un terme dérivé peut 'disparaître' de la combinaison linéaire obtenue par dérivation. C'est pourquoi, pour les expressions à multiplicités, les termes dérivés sont définis par une induction sur des ensembles d'expressions rationnelles. Ce sont les états de l'automate et les combinaisons linéaires obtenues par dérivation permettent d'obtenir les transitions de l'automate des termes dérivés.

Dans ce chapitre nous donnons un algorithme pour construire l'automate des termes dérivés d'une expression rationnelle à multiplicités. Cet algorithme utilise, comme dans le cas booléen, des chemins de nœuds de l'arbre pour représenter les termes dérivés. Un marquage de ces nœuds et des règles de réduction permettent d'identifier les termes dérivés. Les transitions sont également calculées par des calculs sur l'arbre.

Finalement nous proposons une définition pour les termes dérivés cassés. Cette définition introduit le cassage d'une expression à multiplicités sous forme d'une combinaison linéaire d'expressions rationnelles. Les termes dérivés cassés permettant la construction d'un automate, nous donnons également un algorithme pour calculer cet automate. Cet algorithme est une synthèse des résultats pour les termes dérivés cassés booléens et pour les termes dérivés à multiplicités.

5.1 $\mathbb{K}$-expressions rationnelles

Afin de généraliser nos résultats, il nous faut d'abord donner une définition des expressions rationnelles à multiplicités. En particulier il nous faut énoncer les identités triviales qui seront utilisées. Certains choix sont justifiés lors de la définition des termes dérivés.

5.1.1 Définitions et notations

Lorsque l'on traite des séries rationnelles, plus particulièrement de l'étoile de celles-ci, il faut définir une distance, au moins implicitement, sur le semi-anneau des coefficients. Dans tout ce chapitre, $\mathbb{K}$ est donc un semi-anneau topologique, c'est-à-dire que $\mathbb{K}$ est muni d'une topologie pour laquelle l'addition et la multiplication sont continues, et A est un alphabet. L'addition sur $\mathbb{K}$ est notée $\oplus$ et la multiplication est notée par une simple concaténation. Par extension, $\oplus$ note également l'addition dans les séries à coefficients dans $\mathbb{K}$. Le semi-anneau $\mathbb{K}$ n'est, *a priori*, pas commutatif. Sur le semi-anneau topologique $\mathbb{K}$ pour tout élément k, la famille $\{k^n\}_{n\in\mathbb{N}}$ peut être, ou ne pas être, sommable. Lorsqu'elle est sommable, nous appellerons *étoile de k* sa somme, notée k^* :

$$k^* = \sum_{n\in\mathbb{N}} k^n \quad .$$

Lorsqu'elle n'est pas sommable, nous dirons que l'étoile de k n'est pas définie.

Si $\mathbb{K}$ est un semi-anneau topologique *fort*, c'est-à-dire que le produit de deux familles sommables est sommable, alors il est possible de munir le semi-anneau des séries formelles $\mathbb{K}\langle\langle A^*\rangle\rangle$ de la topologie produit dérivée de la topologie de $\mathbb{K}$ – ce qui est plus intéressant que d'utiliser la topologie produit dérivée de la topologie discrète sur $\mathbb{K}$ – (cf. [43, III.1.3]).

Les opérations rationnelles peuvent alors être définies pour les séries rationnelles. Soient s et t deux séries :

$$\begin{aligned} \langle s \oplus t, w\rangle &= \langle s, w\rangle \oplus \langle t, w\rangle \quad , \\ \langle st, w\rangle &= \sigma_{uv=w}\langle s, u\rangle\langle t, v\rangle \quad , \\ s^* &= \sigma_{n\in\mathbb{N}} s^n \quad . \end{aligned}$$

L'étoile de la série s n'est définie que si la série est propre, c'est-à-dire si le coefficient de $1_{\mathbb{K}}$ dans la série s est nul.

Expressions rationnelles à multiplicités

Définition 18. *Une* expression rationnelle sur A^* à multiplicités *dans un semi-anneau* $\mathbb{K}$*, ou* $\mathbb{K}$-expression rationnelle*, est une formule bien parenthésée obtenue inductivement par la grammaire :*

$$\mathsf{E} \to \{a \mid a \in A\} \mid 0 \mid 1 \mid (\mathsf{E}+\mathsf{E}) \mid (\mathsf{E}\cdot\mathsf{E}) \mid (\mathsf{E}^*) \mid (k\mathsf{E}), k \in \mathbb{K} \mid (\mathsf{E}k), k \in \mathbb{K} \quad .$$

On note $\mathbb{K}\,\mathsf{RatE}\,A^*$ l'ensemble des $\mathbb{K}$-expressions rationnelles sur A^*.

L'unité du semi-anneau $\mathbb{K}$ est notée $1_{\mathbb{K}}$ pour éviter toute confusion avec le mot vide 1_{A^*} et l'expression $\mathsf{1}$. Pour éviter également des confusions entre des scalaires et l'expression $\mathsf{1}$ (par exemple dans l'expression (21), nous nous autorisons à noter l'expression $\mathsf{1}$ par 1_{A^*} (qui donne (21_{A^*})) même si nous sommes conscient que le mot vide et l'expression $\mathsf{1}$ ne sont pas les mêmes objets.

Exemple 35. Les formules suivantes sont des expressions rationnelles :

$$((3a) + (2b)) \qquad ((a \cdot ((a^*2) + 1_{A^*}))3) \quad .$$

Remarque 9. Les expressions rationnelles à multiplicités sont souvent (cf. [41] ou [14] par exemple) définies par :

$$\mathsf{E} \to \{a \mid a \in A\} \mid 0 \mid 1 \mid k \in \mathbb{K} \mid (\mathsf{E}+\mathsf{E}) \mid (\mathsf{E}\cdot\mathsf{E}) \mid (\mathsf{E}^*) \quad .$$

C'est-à-dire que les multiplicités sont considérées directement comme des atomes et les multiplications scalaires à gauche et à droite sont alors confondues avec la concaténation. Nous ne prenons pas ce formalisme car dans nos constructions les scalaires et les lettres ne jouent pas le même rôle et les multiplications à droite et à gauche ne sont pas traités comme la concaténation.

Le parenthésage des expressions rationnelles pourra être implicite lorsqu'il n'y a pas de confusion (par exemple les multiplicités à gauche et à droite sont appliquées à la plus petite expression à laquelle elles sont accolées –. De plus, les concaténations à droite consécutives pourront être écrites sans parenthèses.

Nous appelons $\mathbb{K}$-*expression unitaire* (à gauche et à droite) une expression rationnelle qui n'est ni $(k\mathsf{E})$, ni $(\mathsf{E}k)$ – donc soit une somme, soit une concaténation, soit une étoile, soit un atome –.

Exemple 36. L'expression rationnelle $3(a{\cdot}b)$ n'est pas unitaire. En revanche, l'expression $(3a \cdot b)$ l'est (avec le parenthésage explicite, la multiplicité 3 s'applique à a et non à $a \cdot b$).

Comme dans le cas booléen, nous définissons des identités dites *triviales* modulo lesquelles les $\mathbb{K}$-expressions rationnelles sont calculées. Ces identités sont les identités **(T)** du cas booléen auxquelles on ajoute des identités pour les multiplications à gauche et à droite par un scalaire :

$$(0_{\mathbb{K}}\mathsf{E}) \equiv (\mathsf{E}0_{\mathbb{K}}) \equiv 0 \ , \quad (1_{\mathbb{K}}\mathsf{E}) \equiv (\mathsf{E}1_{\mathbb{K}}) \equiv \mathsf{E}$$
$$(1k) \equiv (k1) \ , \quad ((k1) \cdot \mathsf{E}) \equiv (k\mathsf{E}) \ , \quad (\mathsf{E} \cdot (k1)) \equiv (\mathsf{E}k) \ ,$$
$$(k(k'\mathsf{E})) \equiv (kk'\mathsf{E}) \ , \quad ((k\mathsf{E})k') \equiv (k(\mathsf{E}k')) \ , \quad ((\mathsf{E}k)k') \equiv (\mathsf{E}kk') \ .$$

La considération de ces identités permet à nouveau d'éviter d'avoir 0 comme sous-expression d'une expression non nulle. En outre, les trois dernières identités permettent de représenter toute $\mathbb{K}$-expression rationnelle par un triplet (k, E, k') où k et k' sont des éléments de $\mathbb{K}$ et E est une expression unitaire.

Exemple 37. L'expression $2(a \cdot 31_{A^*})3$ se réduit par identités triviales en $2(a9)$.

Remarque 10. Il est courant d'utiliser également comme identité triviale la commutativité des multiplicatition par un scalaire à gauche et à droite : $(ak) \equiv (ka)$. Nous ne la retenons pas car les définitions que nous donnerons des termes dérivés ne commutent pas à cette identité.

Il est également possible de définir une profondeur sur les $\mathbb{K}$-expressions afin de réaliser des raisonnements inductifs sur celles-ci. La *profondeur* d'une expression rationnelle à multiplicités dans $\mathbb{K}$ est définie comme celle d'une expression booléenne, augmenté de :

$$\mathsf{d}(\mathsf{E}k) = \mathsf{d}(k\mathsf{E}) = \mathsf{d}(\mathsf{E}) + 1 \ .$$

Rappelons que l'addition sur le semi-anneau $\mathbb{K}$ est notée $\oplus$, la multiplication implicitement et que les neutres sont $0_{\mathbb{K}}$ et $1_{\mathbb{K}}$.

Définition 19. *Le* terme constant *d'une expression rationnelle est un élément de* $\mathbb{K}$ *défini par :*

$$\begin{gathered}\mathsf{c}(0) = 0_{\mathbb{K}}, \qquad \mathsf{c}(1) = 1_{\mathbb{K}}, \qquad \forall a \in A \quad \mathsf{c}(a) = 0_{\mathbb{K}} \quad , \\ \mathsf{c}((k\mathsf{E})) = k\mathsf{c}(\mathsf{E}), \qquad \mathsf{c}((\mathsf{E}k)) = \mathsf{c}(\mathsf{E})k \quad , \\ \mathsf{c}((\mathsf{F}+\mathsf{G})) = \mathsf{c}(\mathsf{F}) \oplus \mathsf{c}(\mathsf{G}), \qquad \mathsf{c}((\mathsf{F}\cdot\mathsf{G})) = \mathsf{c}(\mathsf{F})\mathsf{c}(\mathsf{G}) \quad , \\ \mathsf{c}((\mathsf{E}^*)) = \mathsf{c}(\mathsf{E})^* \textit{ si le membre de droite est défini.} \quad .\end{gathered}$$

On dit qu'une expression rationnelle est *valide* si son terme constant est défini – et alors toutes ses sous-expressions sont valides –.

Les expressions rationnelles à multiplicités dans $\mathbb{K}$ sont une manière de représenter les séries rationnelles à coefficients dans $\mathbb{K}$. La série *dénotée* par l'expression rationnelle E est définie par :

$$\begin{gathered}|0| = 0 \quad , |1| = 1_{A^*} \quad , |a| = a \quad , \\ |(k\mathsf{E})| = k|\mathsf{E}| \quad , |(\mathsf{E}k)| = |\mathsf{E}|k \quad , \\ |(\mathsf{E}+\mathsf{F})| = |\mathsf{E}| \oplus |\mathsf{F}| \quad , \\ |(\mathsf{E}\cdot\mathsf{F})| = |\mathsf{E}||\mathsf{F}| \quad , \\ |(\mathsf{E}^*)| = |\mathsf{E}|^* \quad .\end{gathered}$$

Comme dans le cas booléen, où le terme constant exprime l'appartenance du mot vide au langage reconnu par une expression rationnelle, le terme constant d'une expression rationnelle à multiplicité est égal à la multiplicité du mot vide 1_{A^*} dans la série rationnelle qu'elle dénote.

Il faut bien faire la différence entre le monôme $k\mathsf{E}$ – que nous pourrons également noter $[k\mathsf{E}]$ pour éviter toute confusion – et l'expression rationnelle $(k\mathsf{E})$. Cette distinction est d'autant plus importante par la suite lorsque les combinaisons linéaires d'expressions sont définies.

L'ensemble des séries rationnelles sur A^* a coefficients dans $\mathbb{K}$ est la clôture rationnelle de l'ensemble des polynômes sur A^* à coefficients dans $\mathbb{K}$.

Proposition 42. *Une série formelle est* rationnelle *si, et seulement si, elle est dénotée par une* $\mathbb{K}$*-expression rationnelle valide.*

Le théorème suivant, relie, comme le théorème de Kleene dans le cas booléen, les séries rationnelles avec les séries reconnaissable (*cf.* [43, Th. III.3.5]).

Théorème 43 (Kleene-Schützenberger). *Une série de $\mathbb{K}\langle\langle A^*\rangle\rangle$ est rationnelle si, et seulement si, elle est le comportement d'un $\mathbb{K}$-automate fini sur A^*.*

Combinaisons linéaires de $\mathbb{K}$-expressions

Par la suite, pour définir les termes dérivés et les termes dérivés cassés, nous avons besoin d'introduire les combinaisons linéaires à gauche d'expressions rationnelles. Ces combinaisons forment un $\mathbb{K}$-module, c'est-à-dire que l'addition est commutative et la multiplication à gauche par un scalaire de $\mathbb{K}$ est distributive :

$$k\mathsf{E} \oplus k'\mathsf{F} = k'\mathsf{F} \oplus k\mathsf{E}, \qquad k\mathsf{E} \oplus k'\mathsf{E} = [k \oplus k']\mathsf{E} \ .$$

Nous ajoutons également à ce module une multiplication à droite par un scalaire et une concaténation à droite par une expression :

$$\begin{aligned} ([k\mathsf{E}] \cdot \mathsf{F}) &\equiv k(\mathsf{E} \cdot \mathsf{F}), & ([k\mathsf{E}]k') &\equiv k(\mathsf{E}k') \ , \\ ([\mathsf{E} \oplus \mathsf{E}'] \cdot \mathsf{F}) &\equiv (\mathsf{E} \cdot \mathsf{F}) \oplus (\mathsf{E}' \cdot \mathsf{F}), & ([\mathsf{E} \oplus \mathsf{E}']k) &\equiv (\mathsf{E}k) \oplus (\mathsf{E}'k) \ . \end{aligned}$$

Nous ne définissons pas une multiplication sur l'ensemble des combinaisons linéaires car, comme la concaténation des expressions rationnelles n'est pas associative, cette multiplication ne serait alors pas associative non plus. L'ensemble des combinaisons linéaires à gauche ne forme donc pas une semi-algèbre.

Arbre syntaxique

Comme nous l'avons déjà fait remarquer, toute $\mathbb{K}$ expression rationnelle peut être représentée par un triplet (k, E, k'), avec E unitaire. Afin de rester proche du modèle booléen, nous choisissons ces représentations en triplets pour *l'arbre syntaxique* d'une $\mathbb{K}$-expression. C'est-à-dire que les nœuds de l'arbre sont des triplets (k, u, k') où u représente un sous-arbre dont la racine est une somme, une concaténation, une étoile ou un atome – *i.e.* un sous-arbre représentant une expression unitaire –. Un tel u est appelé le u-nœud – pour nœud unitaire – du nœud (k, u, k'). Il faut noter que si nous avons choisi cette représentation de l'arbre syntaxique d'une expression pour sa facilité d'écriture et afin de repérer facilement les expressions unitaires, elle a le désavantage de déconnecter la profondeur de l'expression définie précédemment avec celle de l'arbre. La profondeur de l'expression serait celle de l'arbre si nous avions choisi de représenter les multiplications par des scalaires par des nœuds sur l'arbre.

Les *feuilles propres* de l'arbre syntaxique T_{E} d'une expression E sont les u-nœuds de T_{E} étiquetés par une lettre de A. Leur ensemble est toujours noté $\mathrm{fp}(\mathsf{E})$.

Notons $[\![(k,u,k')]\!]$ l'expression rationnelle représentée par le sous-arbre dont la racine est le nœud $x = (k,u,k')$. Notons dans la continuité $[\![u]\!]$ l'expression rationnelle unitaire représentée par u. Ainsi, on a :

$$[\![(k,u,k')]\!] = (k([\![u]\!]k')) \ .$$

Exemple 38. Posons $\mathsf{F}_1 = (2a3 \cdot b)$ et l'expression rationnelle :

$$\mathsf{E}_1 = 3\left((2\mathsf{F}_1 2) + (\mathsf{F}_1^* 2)\right) 3 \ .$$

L'arbre syntaxique de E_1 est montré sur la Figure 5.1. Nous noterons pour la suite u_i le u-nœud associé au nœud x_i. Par exemple u_3 est étiqueté par $*$ et l'expression associée est $[\![u_3]\!] = \mathsf{F}_1^*$.

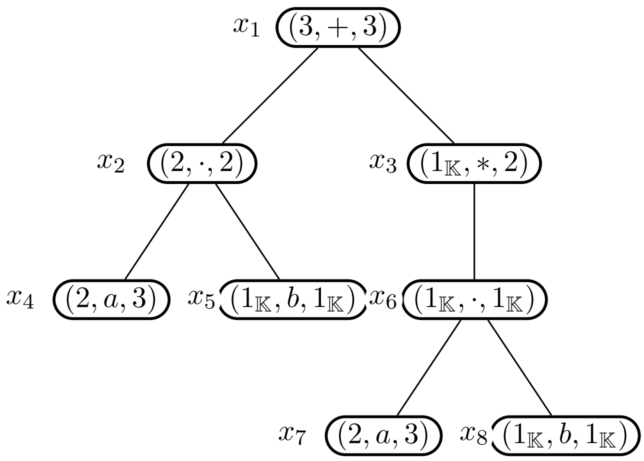

FIGURE 5.1 – L'arbre syntaxique de E_1

Un u-*nœud pondéré* est un triplet (k, u, k') où k et k' sont des éléments de $\mathbb{K}$. Il est important de noter que si un nœud x de l'arbre est bien un u-nœud pondéré, la réciproque n'est pas vraie. C'est-à-dire que nous nous autorisons à donner des coefficients aux u-nœuds qui n'apparaissent pas nécessairement dans l'arbre syntaxique de l'expression – seule l'expression unitaire doit apparaître dans l'arbre –.

Sur les u-nœuds pondérés il est possible d'étendre la définition d'expression rationnelle associée et de définir un terme constant naturellement :

(i) $[\![(k, u, k')]\!] = (k([\![u]\!]k'))$,

(ii) $\mathsf{c}((k, u, k')) = \mathsf{c}([\![(k, u, k')]\!])$.

Exemple 39 (*Ex. 38 cont.*)**.** Par exemple le u-nœud pondéré $(2, u_6, 2)$ n'est pas un nœud de l'arbre $\mathrm{T}_{\mathsf{E}_1}$. Il représente pourtant la même expression que le nœud x_2 : $[\![(2, u_6, 2)]\!] = 2\mathsf{F}_1 2$.

Nous étendons également la notion de premières feuilles d'un nœud – ou d'une expression – pour les expressions à multiplicités :

Définition 20. *La fonction* First *associe à tout u-nœud pondéré de l'arbre syntaxique* T_{E} *de l'expression* E *une combinaison linéaire à gauche de feuilles propres définie inductivement par :*

$$\begin{aligned}
\mathsf{First}(0) = \mathsf{First}(1) &= 0_{\mathbb{K}} \quad , \\
\mathsf{First}((k, u, k')) &= \begin{cases} ku & \text{si } u \in \mathrm{fp}(\mathsf{E}) \\ k\mathsf{First}(u) & \text{sinon.} \end{cases} \quad , \\
\mathsf{First}(x + y) &= \mathsf{First}(x) \oplus \mathsf{First}(y) \quad , \\
\mathsf{First}(x \cdot y) &= \mathsf{First}(x) \oplus \mathsf{c}(x)\mathsf{First}(y) \quad , \\
\mathsf{First}(x^*) &= \mathsf{c}(x)^*\mathsf{First}(x) \quad ,
\end{aligned}$$

où (k, u, k')*,* x *et* y *sont des u-nœuds pondérés.*

Exemple 40 (*Ex. 38 cont.*)**.** On a par exemple :

$$\mathsf{First}(x_1) = 12u_4 \oplus 6u_7 \quad .$$

Comme nous le verrons par la suite, les termes dérivés et les termes dérivés cassés d'expressions à multiplicités sont des concaténations à droite pondérées de sous-expressions unitaires. C'est pourquoi nous considérons les mots dont les lettres sont des u-nœuds pondérés. Le mot ε représente le mot vide sur cet alphabet. Sur ces mots, à l'instar des mots de nœuds booléens, nous étendons les fonctions terme constant et First :

(i) $\mathsf{c}(x_1 :: x_2 :: \ldots :: x_n) = \mathsf{c}(x_1)\mathsf{c}(x_2)\ldots\mathsf{c}(x_n)$,
(ii) $\mathsf{First}(\varepsilon) = 0_{\mathbb{K}}$,
(iii) $\mathsf{First}(x_1 :: \ldots :: x_n) = \mathsf{First}(x_1 :: \ldots :: x_{n-1}) \oplus \mathsf{c}(x_1 :: \ldots :: x_{n-1})\mathsf{First}(x_n)$.

Pour simplifier les écritures, nous nous autorisons à simplifier la notation des u-nœuds pondérés dont l'expression unitaire est 1. Ces expressions apparaîtront beaucoup dans les mots que nous calculerons et, par le jeu des expressions triviales, représentent sur l'expression dénotée par un mot, non pas une concaténation, mais une multiplication à droite par un scalaire. Ainsi, nous simplifions le u-nœud pondéré $(k', 1, k'')$ en k, avec $k = k'k''$. Un tel u-nœud pondéré sera naturellement appelé une multiplicité, ou *poids*.

Exemple 41 (*Ex. 38 cont.*)**.** Par exemple on a :

$$\begin{aligned}
[\![3 :: x_5 :: 2 :: 3]\!] &= (3(b6)) \\
[\![3 :: x_8 :: (1_{\mathbb{K}}, x_3, 1_{\mathbb{K}}) :: 2 :: 3]\!] &= ((3b) \cdot \mathsf{F}_1^*)6 \\
[\![2 :: 3]\!] &= (61_{A^*}) \\
[\![x_3 :: 2 :: 3]\!] &= \mathsf{F}_1^*6 \quad .
\end{aligned}$$

Et on a :

$$\begin{aligned}
\mathsf{First}(3 :: x_5 :: 2 :: 3) &= 3u_5 \\
\mathsf{First}(3 :: x_8 :: (1_{\mathbb{K}}, x_3, 1_{\mathbb{K}}) :: 2 :: 3) &= 3u_8 \\
\mathsf{First}(2 :: 3) &= 0 \\
\mathsf{First}(x_3 :: 2 :: 3) &= 2u_7 \quad .
\end{aligned}$$

5.1.2 Représentation canonique

Dans le cas booléen, les termes dérivés étaient représentés par des mots sur les nœuds car ils étaient des concaténations à droite de sous-expressions. Pour les expressions à multiplicités, nous allons montrer par la suite que les termes dérivés, et les termes dérivés cassés, sont des concaténations de sous-expressions unitaires pondérées. Il apparaît donc que les termes dérivés pourront être représentés par des mots sur les u-nœuds pondérés. Nous allons donc développer dans cette sous-section une méthode similaire à celle du cas booléen, avec des marques sur les nœuds et u-nœuds, qui permet de les identifier et d'avoir une représentation canonique d'une concaténation de sous-expressions unitaires pondérées par un mot sur les u-nœuds pondérés. Plusieurs particularités des multiplicités rendent la tâche plus complexe que dans le cas booléen.

Dans le cas booléen, il était possible de ne s'intéresser qu'aux mots propres car il suffisait 'd'effacer' les lettres 1 du mot sans que cela ne change l'expression associée. Ceci n'est pas aussi direct dans le cas à multiplicités car seules les lettres $(1_{\mathbb{K}}, 1, 1_{\mathbb{K}})$ peuvent être effacées. Les autres poids non égaux à l'unité doivent être traités différemment. En particulier il est nécessaire, avant de faire tout autre calcul, de transformer le mot afin de se 'conformer' aux identités triviales introduites par les multiplicités. La première opération est alors de regrouper les poids successifs.

Les marques

L'arbre syntaxique de l'expression E est 'marqué' de telle manière que chaque u-nœud u a une marque $\overline{u}$ telle que $\overline{u} = \overline{v}$ si $\|u\| = \|v\|$, c'est-à-dire que u et v sont racines d'arbres isomorphes. La définition de marque s'étend aux u-nœuds pondérés par :

$$\overline{(k, u, k')} = (k, \overline{u}, k') \ .$$

Dans un souci de simplicité, nous continuons d'appeler *poids* les marques des poids (les u-nœuds pondérés dont l'expression unitaire est l'unité). Les mots sur les marques sont les mots dont les lettres sont des marques de u-nœuds pondérés.

Exemple 42 (*Ex. 38 cont.*)**.** Notons α_i la marque du u-nœud u_i. On a alors les égalités suivantes :

$$\alpha_4 = \alpha_7 \qquad \alpha_5 = \alpha_8$$

de plus comme les poids sont identiques, on a $\overline{x_4} = \overline{x_7}$ et $\overline{x_5} = \overline{x_8}$ et donc on a également :

$$\alpha_2 = \alpha_6 \ .$$

Il est important de remarquer que l'on n'a néanmoins pas $\overline{x_2} = \overline{x_6}$ car les poids diffèrent.

Les identités triviales

Dans un premier temps, pour transformer un mot sur les marques en un mot canonique pour l'expression associée, nous allons le transformer de la même manière que les identités triviales permettent de transformer des expressions. Il faut noter que, contrairement aux réductions vues dans le cas booléen – et dont l'adaptation aux multiplicités suit –, les réductions que nous effectuons ici sont totalement indépendantes de l'expression E.

Nous allons ainsi tout d'abord transformer un mot sur les marques afin qu'il soit cohérent avec les identités :

$$((\mathsf{E}k)k') = (\mathsf{E}kk') \qquad \text{et} \quad (k(k'\mathsf{E})) = (kk'\mathsf{E}) \ .$$

Pour cela il suffit de remplacer tous poids consécutifs en leur produit :

$$\alpha_1 :: \ldots :: k :: k' :: \cdots \rightarrow^{\rho_1} \alpha_1 :: \ldots :: kk' :: \ldots \ .$$

Exemple 43 (*Ex. 38 cont.*)**.** Par exemple $3 :: \overline{x_5} :: 2 :: 3$ se réduit par ρ_1 en $3 :: \overline{x_5} :: 6$.

Dans un deuxième temps, nous rendons cohérent le mot pour l'identité :

$$((k\mathsf{1}) \cdot (k'\mathsf{E})) = (kk'\mathsf{E}) \ .$$

Pour cela il faut rentrer le coefficient dans le premier terme de la concaténation à droite :

$$k :: (k_1, \overline{u_1}, k'_1) :: \cdots \rightarrow^{\rho_2} (kk_1, \overline{u_1}, k'_1) :: \ldots \ .$$

Exemple 44 (*Ex. 38 cont.*)**.** Par exemple $3 :: \overline{x_5} :: 6$ se réduit par ρ_2 en $(3, \alpha_5, 1_{\mathbb{K}}) :: 6$.

Finalement, nous transformons l'expression pour la rendre cohérente avec :

$$((\mathsf{E}k) \cdot (k'\mathsf{1}) = (\mathsf{E}kk')$$

Comme les mots sur les marques représentent des concaténations à droite, le coefficient en fin de mot ne doit être rentré dans la marque précédente que dans le cas où celle-ci est également la première marque du mot :

$$(k_1, \overline{u_1}, k'_1) :: k \rightarrow^{\rho_3} (k_1, \overline{u_1}, k'_1 k) \ .$$

Exemple 45 (*Ex. 38 cont.*)**.** Par exemple $(3, \alpha_5, 1_{\mathbb{K}}) :: 6$ se réduit par ρ_3 en $(3, \alpha_5, 6)$. En revanche, $(3, \alpha_5, 1_{\mathbb{K}}) :: 6 :: (2, \alpha_1, 3)$ ne se réduit pas par ρ_3.

Après avoir réalisé successivement les transformations ρ_1, ρ_2 et ρ_3, un mot de marque $\overline{X}$ quelconque est donc remplacé par un mot de marque $\overline{Y}i = \rho_3(\rho_2(\rho_1(\overline{X})))$ tel que :

(i) il n'y a pas deux poids consécutifs dans $\overline{Y}$;
(ii) $\overline{Y}$ ne commence pas par un poids ;
(iii) si $\overline{Y}$ est de longueur 2 alors ses lettres ne sont pas des poids ;
(iv) $\overline{X}$ et $\overline{Y}$ représentent la même expression.

Réduction

Lorsque le mot a été transformé pour être cohérent avec les identités triviales, nous allons réaliser les réductions qui permettent d'avoir une représentation canonique des concaténations de sous-expressions. Contrairement aux transformations précédentes, les règles de réductions présentées maintenant sont dépendantes de l'expression rationnelle, ou plus particulièrement de son arbre syntaxique, que l'on considère.

Soit $\mathcal{R}$ l'ensemble des règles de réécriture sur les mots de marques :

$$(k_1, \alpha_1, k_1') :: (k_2, \alpha_2, k_2') \to (1_{\mathbb{K}}, \overline{u}, 1_{\mathbb{K}})$$

avec α_1, α_2 et u tels qu'il existe $v = x_1 \cdot x_2$ où $x_1 = (k_1, u_1, k_1')$ et $x_2 = (k_2, u_2, k_2')$ et avec $\overline{u_1} = \alpha_1$, $\overline{u_2} = \alpha_2$ et $\overline{v} = \overline{u}$. C'est-à-dire des nœuds dont les marques sont respectivement α_1 et α_2 et qui sont fils gauche et droit d'un u-nœud étiqueté par une concaténation et dont la marque est $\overline{u}$.

Exemple 46 (*Ex. 38 cont.*)**.** Pour E_1 il y a une seule règle dans $\mathcal{R}$ car les deux concaténations amènent la même règle sur les marques :

$$(2, \alpha_4, 3) :: (1_{\mathbb{K}}, \alpha_5, 1_{\mathbb{K}}) \to (1_{\mathbb{K}}, \alpha_2, 1_{\mathbb{K}}) \quad .$$

Notons ρ_4 la réduction préfixe par l'ensemble des règles $\mathcal{R}$. La réduction complète est notée ρ : pour tout mot sur les marques $\overline{X}$, on a $\rho\overline{X} = \rho_3(\rho_4(\rho_3(\rho_2(\rho_1(\overline{X})))))$. Le dernière occurrence de ρ_3 vient du fait que la réduction préfixe ρ_4 réduit la taille du mot et il est possible qu'il devienne de longueur 2 et il faut alors à nouveau vérifier si la deuxième marque est un poids. Avec cet ajout, il apparaît que la réécriture ρ est idempotente. En outre toutes les transformations ont été choisies dans le souci de conserver l'expression rationnelle dénotée :

$$\|\rho\overline{X}\| = \|\overline{X}\| \quad .$$

La réduction ρ permet l'écriture de manière canonique de concaténations à droite de sous-expressions unitaires pondérées comme le prouve le théorème suivant :

Théorème 44. *Soient $\overline{X}$ et $\overline{Y}$ deux mots sur les marques. On a :*

$$\|\overline{X}\| = \|\overline{Y}\| \Leftrightarrow \rho\left(\overline{X}\right) = \rho\left(\overline{Y}\right) \quad .$$

Démonstration. Comme la réduction est idempotente, nous allons montrer que si $\overline{X}$ et $\overline{Y}$ sont réduits par ρ, alors ils sont égaux si, et seulement si, ils dénotent la même expression. La réciproque est triviale, montrons donc que s'ils dénotent la même expression, alors ils sont égaux. Soient $\overline{X} = \overline{x_1} :: \ldots :: \overline{x_n}$ et $\overline{Y} = \overline{y_1} :: \ldots :: \overline{y_m}$ deux mots réduits sur les marques, avec $x_i = (k_i, u_i, k'_i)$ et $y_i = (h_i, v_i, h'_i)$.

Par induction sur n et sur m : **Cas** $n = 1$:

$m = 1$: $\|\overline{X}\| = \|\overline{Y}\|$ si, et seulement si, $\|\overline{x_1}\| = \|\overline{y_1}\|$, c'est-à-dire $\|u_1\| = \|v_1\|$, $k_1 = h_1$ et $k'_1 = h'_1$, soit $\overline{X} = \overline{Y}$.

$m = 2$: comme $\overline{Y}$ est réduit, ni y_1, ni y_2 ne sont des poids. Dans ce cas $\|\overline{X}\| = \|\overline{Y}\|$ si, et seulement si, $\|\overline{x_1}\| = \|\overline{y_1}\| \cdot \|\overline{y_2}\|$. Ce qui implique $k_1 = k'_1 = 1_{\mathbb{K}}$ et il existe x tel que $\overline{x} = \overline{x_1}$ et $x = y_1 \cdot y_2$. Ce cas est impossible puisque $\overline{Y}$ a été réduit par ρ_4.

$m > 2$: Ce cas se réduit inductivement au cas précédent qui est impossible.

Cas $n > 1$ **et** $m > 1$: par identification sur l'expression rationnelle $\|\overline{X}\|$, on a $\|\overline{X}\| = \|\overline{Y}\|$ seulement si $\overline{x_n} = \overline{y_m}$ et l'on peut donc revenir au premier cas. □

5.2 Termes dérivés

Dans cette section nous rappelons la définition des termes dérivés et de la dérivation qui leur est associée dans le cas des expressions rationnelles à multiplicités. Cette définition a été donnée dans [31]. Par rapport au cas booléen, un certain nombre de choix sont à faire si l'on souhaite étendre la définition des termes dérivés, en particulier en ce qui concerne le traitement des expressions avec des poids à gauche et à droite. La définition qui a été retenue l'a été afin d'avoir, comme dans le cas booléen, l'automate des termes dérivés quotient de l'automate des positions. Nous verrons entre autre que cette condition amène à considérer différemment les multiplications à gauche et à droite, ce qui justifie notre choix de ne pas construire les expressions modulo la commutativité des multiplicités sur les atomes.

5.2.1 Définition

Définition 21 ([31])**.** *Soit* E *une expression* $\mathbb{K}$*-rationnelle et* a *une lettre de* A*. La* $\mathbb{K}$-derivation *de* E *par rapport à* a*, notée* $\dfrac{\partial}{\partial a}\mathsf{E}$*, est la combinaison linéaire à gauche d'expressions rationnelles à coefficients dans* $\mathbb{K}$*, définie par* *les formules suivantes :*

$$\frac{\partial}{\partial a}\mathsf{0} = \frac{\partial}{\partial a}\mathsf{1} = \mathsf{0}\,, \quad \frac{\partial}{\partial a} b = \begin{cases} \mathsf{1} & si \quad b = a \\ \mathsf{0} & sinon \end{cases} \quad ,$$

$$\frac{\partial}{\partial a}(k\,\mathsf{E}) = k\,\frac{\partial}{\partial a}\,\mathsf{E}\,, \quad \frac{\partial}{\partial a}(\mathsf{E}\,k) = \left(\left[\frac{\partial}{\partial a}\,\mathsf{E}\right]\,k\right) \quad ,$$

$$\frac{\partial}{\partial a}(\mathsf{E}{+}\mathsf{F}) = \frac{\partial}{\partial a}\,\mathsf{E} \oplus \frac{\partial}{\partial a}\,\mathsf{F} \quad , \tag{5.1}$$

$$\frac{\partial}{\partial a}(\mathsf{E}\cdot\mathsf{F}) = \left(\left[\frac{\partial}{\partial a}\,\mathsf{E}\right]\cdot\mathsf{F}\right) \oplus \mathsf{c}(\mathsf{E})\,\frac{\partial}{\partial a}\,\mathsf{F} \quad , \tag{5.2}$$

$$\frac{\partial}{\partial a}(\mathsf{E}^*) = \mathsf{c}(\mathsf{E})^*\,\left(\left[\frac{\partial}{\partial a}\,\mathsf{E}\right]\cdot(\mathsf{E}^*)\right) \quad . \tag{5.3}$$

La dérivation d'une combinaison linéaire est définie par linéarité :

$$\frac{\partial}{\partial a}\left(\bigoplus_{i\in I} k_i\,\mathsf{E}_i\right) = \bigoplus_{i\in I} k_i\,\frac{\partial}{\partial a}\,\mathsf{E}_i \quad . \tag{5.4}$$

Exemple 47 (*Ex. 38 cont.*)**.** La dérivation de E_1 par rapport à a donne :

$$\begin{aligned}
\frac{\partial}{\partial a}\,\mathsf{F}_1 &= 2(\frac{\partial}{\partial a}(a3)\cdot b) \\
&= 2((1_{A^*}3)\cdot b) = [2(3b)] \\
\frac{\partial}{\partial a}\,\mathsf{E}_1 &= 3\left(\frac{\partial}{\partial a}((2\mathsf{F}_1 2) + (\mathsf{F}_1^* 2))\right)3 \\
&= 3\left(2(\frac{\partial}{\partial a}\,\mathsf{F}_1 2) \oplus ((\frac{\partial}{\partial a}\,\mathsf{F}_1\cdot\mathsf{F}_1^*)2\right)3 \\
&= [12(3b2)] \oplus [6(((3b)\cdot\mathsf{F}_1^*)6)] \quad .
\end{aligned}$$

Nous rappelons que $[12(3b2)]$ est le monôme de coefficient 12 et d'expression rationnelle $(3b2)$, ce qui est différent de $[36(b2)]$ qui est le monôme de coefficient 36 de l'expression $(b2)$.

Cette dérivation par rapport à une lettre permet une dérivation par rapport à un mot :

Définition 22. *La dérivation de* E *par rapport à un mot non vide* ua *de* A^* *est définie par induction par :*

$$\frac{\partial}{\partial ua}\,\mathsf{E} = \frac{\partial}{\partial a}\left(\frac{\partial}{\partial u}\,\mathsf{E}\right) \quad .$$

Exemple 48 (*Ex. 38 cont.*)**.** La dérivation de E_1 par rapport à ab et aba donnent :

$$\begin{aligned}\frac{\partial}{\partial ab}\,\mathsf{E}_1 &= \frac{\partial}{\partial b}\,12(3b2) \oplus 6(((3b)\cdot \mathsf{F}_1)6)\\ &= 36(21_{A^*}) \oplus 18(\mathsf{F}_1 6)\\ \frac{\partial}{\partial aba}\,\mathsf{E}_1 &= \frac{\partial}{\partial a}\,36(21_{A^*}) \oplus 18(\mathsf{F}_1 6)\\ &= 36((3b)6)\quad .\end{aligned}$$

Contrairement au cas booléen, la définition de termes dérivés à partir des termes qui apparaissent dans les dérivations par rapport à un mot et la définition inductive de l'ensemble des termes dérivés ne donnent pas le même résultat. Dans le but de garder, dans le cas à multiplicités, l'automate des termes dérivés comme quotient de l'automate des positions de l'expression, nous prenons la définition inductive des termes dérivés :

Définition 23 ([31])**.** *L'ensemble* $\mathrm{TD}(\mathsf{E})$ *des* vrais termes dérivés *de l'expression* E *est défini inductivement par les règles suivantes :*

$$\begin{gathered}\mathrm{TD}(\mathsf{0}) = \mathrm{TD}(\mathsf{1}) = \emptyset\quad ,\quad \forall a \in A,\ \mathrm{TD}(a) = \{\mathsf{1}\}\quad ,\\ \forall k \in \mathbb{K},\ \mathrm{TD}(k\,\mathsf{E}) = \mathrm{TD}(\mathsf{E})\quad ,\quad \mathrm{TD}(\mathsf{E}\,k) = (\mathrm{TD}(\mathsf{E})\ k)\quad ,\\ \mathrm{TD}(\mathsf{E}+\mathsf{F}) = \mathrm{TD}(\mathsf{E}) \ \cup\ \mathrm{TD}(\mathsf{F})\quad ,\\ \mathrm{TD}(\mathsf{E}\cdot\mathsf{F}) = (\mathrm{TD}(\mathsf{E})\cdot\mathsf{F}) \ \cup\ \mathrm{TD}(\mathsf{F})\quad ,\\ \mathrm{TD}(\mathsf{E}^*) = (\mathrm{TD}(\mathsf{E})\cdot(\mathsf{E}^*))\quad .\end{gathered}$$

L'expression E n'est pas nécéssairement dans $\mathrm{TD}(\mathsf{E})$ et l'ensemble des *termes dérivés* de E est : $\mathrm{D}(\mathsf{E}) = \mathrm{TD}(\mathsf{E}) \cup \{\mathsf{E}\}$.

Exemple 49 (*Ex. 38 cont.*)**.** Les vrais termes dérivés de E_1 sont :

$$\begin{aligned}\mathrm{TD}(\mathsf{F}_1) &= \mathrm{TD}((a3)\cdot b)\\ &= \{(1_{A^*}3)\cdot b\} \cup \mathrm{TD}(b) = \{(3b), 1_{A^*}\}\\ \mathrm{TD}(\mathsf{E}_1) &\\ &= \mathrm{TD}((2\mathsf{F}_1 2) + (\mathsf{F}_1^* 2))\,3\\ &= (\mathrm{TD}(\mathsf{F}_1 2)\,3)\ \cup\ (\mathrm{TD}(\mathsf{F}_1^* 2)\,3)\\ &= \{3b6, 6, ((3b)\cdot\mathsf{F}_1^*)6, (\mathsf{F}_1^* 6)\}\quad .\end{aligned}$$

La définition inductive des vrais termes dérivés – et celle des termes dérivés qui en découle – permet d'énoncer la proposition suivante :

Proposition 45. *Les termes dérivés d'une expressions rationnelle* E *sont des concaténations à droite de sous-expressions unitaires de* E *pondérées.*

Une induction sur la profondeur de l'expression permet de relier la dérivation avec l'ensemble des termes dérivés :

Théorème 46 ([31])**.** *Soit* $\mathrm{D}(\mathsf{E}) = \{\mathsf{K}_1, ..., \mathsf{K}_n\}$ *l'ensemble des termes dérivés de* E*. Pour toute lettre* a *de* A*, il existe une matrice carrée* $a\mu$*, de dimension* n *et à coefficients dans* $\mathbb{K}$ *telle que :*

$$\forall i \leqslant n \qquad \frac{\partial}{\partial a}\mathsf{K}_i = \bigoplus_j a\mu_{i,j}\,\mathsf{K}_j\ .$$

Définition 24. L'automate des termes dérivés *d'une* $\mathbb{K}$*-expression* E *est l'automate* $\mathcal{D}(\mathsf{E})$ *dont les états sont les termes dérivés.*
(i) *La fonction initiale vaut* 1 *pour* E *et* 0 *pour tous les autres termes dérives.*
(ii) *La fonction finale est le terme constant du terme dérivé.*
(iii) *Il y a une transition de* K_i *à* K_j *étiqueté par* ka *si* $k = a\mu_{i,j} \neq 0_{\mathbb{K}}$*.*

Exemple 50 (*Ex. 38 cont.*)**.** La Figure 5.2 représente l'automate des termes dérivés de E_1.

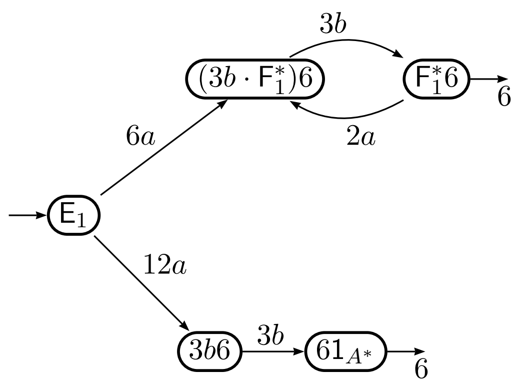

FIGURE 5.2 – L'automate $\mathcal{D}(\mathsf{E}_1)$

L'automate $\mathcal{D}(\mathsf{E})$ reconnaît la série dénotée par E et est un quotient de l'automate des positions [31]. C'est ce résultat qui a motivé deux choix dans nos définitions qui sont explicités dans les deux remarques suivantes :

Remarque 11 ([31])**.** Nous avons choisi comme définition pour les termes dérivés avec multiplicités la définition inductive. Ce n'est pas, comme la généralisation du cas booléen pourrait faire penser, l'ensemble des expressions qui apparaissent avec un coefficient non nul dans une dérivation par rapport à un mot non vide.

Ceci s'explique par la possibilité qu'un terme apparaissant lors d'une dérivation par rapport à une lettre d'un terme dérivé n'apparaisse pas dans la dérivation par rapport à un mot non-vide car il peut exister plusieurs chemins dont les coefficients s'annulent. Il est même possible que la dérivation par rapport à une lettre – et non un mot – ne fasse pas apparaître certains termes dérivés.

Par exemple, si l'on prend l'expression rationnelle à coefficients dans $\mathbb{Z}$: $a + (-1a)$ alors la dérivation par rapport à a donne :

$$\frac{\partial}{\partial a}\, a + (-1)a = 1 \oplus -1[1] = 0 \ \ .$$

Les dérivations par rapport à tout mot non vide donnent le même résultat. Si l'on prenait une définition des termes dérivés basés sur la dérivation, l'expression $a+(-1a)$ aurait pour automate de termes dérivés un automate avec un unique état initial. Ce n'est pas le quotient de l'automate des positions qui a un état initial et un état final (sans transition).

La définition des termes dérivés inductive donne $\mathrm{TD}\,(a + (-1a)) = \{1\}$, ce qui correspond donc a un automate des termes dérivés à deux états, un état initial, un état final et aucune transition. Cet automate est bien un quotient de l'automate des positions, qui a trois états, un initial, deux finaux, et des transitions étiquetées respectivement a et$-a$ de l'état initial aux deux états finaux (*cf.*Figure 5.4).

Figure 5.3 – L'automate $\mathcal{D}(a + (-1)a)$

Remarque 12. De la même manière, le choix de ne pas calculer les expressions modulo l'identité $(ak) \equiv (ka)$ se justifie par le fait que l'on veut que l'automate des termes dérivés soit un quotient de l'automate des positions.

En effet, l'expression rationnelle $((c \cdot b)k)$ est telle qu'avec cette identité, l'automate des termes dérivés construit ne serait pas le quotient de l'automate des positions. Dans ce cas c'est même le Théorème 46 qui ne s'applique

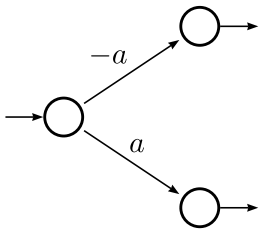

FIGURE 5.4 – L'automate des positions de $a + (-1)a$

plus, en effet :

$$\frac{\partial}{\partial c}(c \cdot b)k = (bk) \equiv (kb)$$
$$\frac{\partial}{\partial b} kb = k1_{A^*} \quad .$$

Pourtant la définition inductive donne :

$$\mathrm{TD}\left((c \cdot b)k\right) = \mathrm{TD}\left(c \cdot b\right) k = \{(bk), (1k)\} \equiv \{(kb), (k1)\}.$$

Le terme dérivé 1 n'apparaît pas dans la définition inductive, c'est $k1$ qui le remplace.

5.2.2 Calcul de l'automate

Chemins et termes dérivés

Afin de construire l'automate des termes dérivés d'une expression à multiplicités dans $\mathbb{K}$, nous allons d'abord montrer comment, comme dans le cas booléen, les termes dérivés peuvent être exprimés sous formes de mots sur les u-nœuds pondérés obtenus par un chemin dans l'arbre syntaxique.

Il est toujours possible de décorer l'arbre syntaxique à la manière de ce qui est fait dans le cas booléen. Seule une différence pour les liens : le lien qui part d'un fils gauche d'un nœud concaténation pointe sur son frère, comme dans le cas booléen, mais le lien qui part d'un fils d'un nœud étoilé pointe vers le u-nœud de son père. Cette modification se justifie par le fait que si deux u-nœuds sont fils d'une concaténation, alors l'expression qui est représentée est bien la concaténation des deux u-nœuds. Par contre, lorsqu'une sous-expression est étoilée, elle est unitaire et donc indépendante des poids sur le nœud père.

Les chemins qui permettent d'obtenir les termes dérivés sont les mêmes physiquement sur l'arbre que dans le cas booléen mais le mot qui leur est associé est différent car les poids droits sont récoltés également tout le long du trajet.

Formellement, *le chemin associé* au u-nœud pondéré $x = (k, u, k')$ est défini par le mot suivant :

$$\pi_r(x) = \begin{cases} \varepsilon & \text{si } x \text{ est la racine et si } k' = 1_{\mathbb{K}}, \\ k' & \text{si } x \text{ est la racine et } k' \neq 1_{\mathbb{K}}, \\ k' :: (1_{\mathbb{K}}, v, 1_{\mathbb{K}}) :: \pi_r(\phi(x)) & \text{si } v = x^*, \\ k' :: y :: \pi_r(\phi(x)) & \text{si } \phi(x) = x \cdot y, \\ k' :: \pi_r(\phi(x)) & \text{sinon.} \end{cases}$$

Lorsqu'il n'y a pas d'ambiguïté sur la racine, il est possible d'écrire $\pi(x)$. De même, si u est un u-nœud, il sera possible d'écrire $\pi(u)$ pour $\pi((1_{\mathbb{K}}, u, 1_{\mathbb{K}}))$.

Exemple 51 (*Ex. 38 cont.*)**.** On a par exemple :

$$\begin{aligned} \pi(u_4) &= 3 :: x_5 :: 2 :: 3 \ \ , \\ \pi(u_5) &= 2 :: 3 \ \ , \\ \pi(u_7) &= 3 :: x_8 :: (1_{\mathbb{K}}, u_3, 1_{\mathbb{K}}) :: 2 :: 3 \ \ , \\ \pi(u_8) &= (1_{\mathbb{K}}, u_3, 1_{\mathbb{K}}) :: 2 :: 3 \ \ . \end{aligned}$$

Comme dans le cas booléen, on se rend compte sur cet exemple, à travers des calculs déjà réalisés, que les expressions associées aux chemins des feuilles sont les termes dérivés.

Théorème 47. *Soit* E *une expression rationnelle,* T_{E} *son arbre syntaxique de racine* r*. Si* $\mathsf{First}(r, a) = \bigoplus_i k_i \ell_i$*, alors :*

$$\frac{\partial}{\partial a} \mathsf{E} = \bigoplus_i k_i [\![\pi(\ell_i)]\!] \ \ .$$

Démonstration. Dans cette preuve, nous allons calculer des chemins dans des sous arbres, nous allons donc préciser la racine de l'arbre dans lequel le chemin est calculé. Lorsque la racine sera omise, nous seront dans l'arbre complet, de racine r. La preuve de ce théorème est une induction sur la profondeur de E. Les cas de base où E est un atome sont triviaux. Supposons donc que $r = (k, u, k')$.

– Si $u = x + y$, c'est-à-dire que $\mathsf{E} = (k((\mathsf{F} + \mathsf{G})k'))$, alors :

$$\mathsf{First}(\mathsf{E}, a) = k\mathsf{First}(u, a) = k(\mathsf{First}(x, a) \oplus \mathsf{First}(y, a)) \ \ ,$$

Si $\mathsf{First}(x,a) = \bigoplus_j k_j\ell_j$, avec $k_j \neq 0_{\mathbb{K}}$, alors on a :

$$\pi_r(\ell_j) = \pi_x(\ell_j) :: k' \quad ,$$

et donc $\llbracket \pi_r(\ell_j) \rrbracket = (\llbracket \pi_x(\ell_j) \rrbracket k')$.
De la même manière pour les feuilles de G : si $\mathsf{First}(y,a) = \bigoplus_h k_h\ell_h$ alors $\llbracket \pi_r(\ell_h) \rrbracket = (\llbracket \pi_y(\ell_h) \rrbracket k')$.
Finalement :

$$\begin{aligned}
\bigoplus_i k_i \llbracket \pi(\ell_i) \rrbracket &= k((\bigoplus_j k_j \llbracket \pi(\ell_j) \rrbracket k') \oplus (\bigoplus_h k_h \llbracket \pi(\ell_h) \rrbracket k')) \\
&= k\left(\frac{\partial}{\partial a}(\mathsf{F}k' + \mathsf{G}k')\right) \\
&= \frac{\partial}{\partial a}\mathsf{E} \quad .
\end{aligned}$$

– Si $u = x \cdot y$, c'est-à-dire $\mathsf{E} = (k((\mathsf{F} \cdot \mathsf{G})k'))$, alors :

$$\mathsf{First}(\mathsf{E},a) = k\mathsf{First}(u,a) = k(\mathsf{First}(x,a) \oplus \mathsf{c}(x)\mathsf{First}(y,a)) \quad .$$

Si $y = (h, v, h')$ et si $\mathsf{First}(x,a) = \bigoplus_j k_j\ell_j$, avec $k_j \neq 0_{\mathbb{K}}$, alors on a :

$$\pi_u(\ell_j) = \pi_x(\ell_j) :: (h, v, 1_{\mathbb{K}}) :: h' :: k' \quad ,$$

donc $\llbracket \pi_u(\ell_j) \rrbracket = ((\llbracket \pi_x(\ell_j) \rrbracket \cdot \mathsf{G})k')$.
Si $\mathsf{First}(y,a) = \bigoplus_h k_h\ell_h$ alors $\llbracket \pi_r(l_h) \rrbracket = (\llbracket \pi_y(l_h) \rrbracket k')$ et donc :

$$\begin{aligned}
\bigoplus_i k_i \llbracket \pi(\ell_i) \rrbracket &= k(\bigoplus_j k_j((\llbracket \pi(\ell_j) \rrbracket \cdot \mathsf{G})k') \bigoplus_h \mathsf{c}(s) k_h(\llbracket \pi(\ell_h) \rrbracket k')) \\
&= k\left(\frac{\partial}{\partial a}(\mathsf{F} \cdot \mathsf{G})k'\right) \\
&= \frac{\partial}{\partial a}\mathsf{E} \quad .
\end{aligned}$$

– Si $u = x^*$, c'est-à-dire $\mathsf{E} = (k((\mathsf{F}^*)k'))$, alors :

$$\mathsf{First}(\mathsf{E},a) = k\mathsf{c}(x)^*\mathsf{First}(x,a) \quad .$$

Si $\mathsf{First}(x,a) = \bigoplus_j k_j\ell_j$, avec $k_j \neq 0_{\mathbb{K}}$, alors on a :

$$\pi_r(\ell_j) = \pi_u(\ell_j) :: (1_{\mathbb{K}}, x, 1_{\mathbb{K}}) :: k' \quad .$$

donc $\llbracket \pi_r(\ell_j) \rrbracket = ((\llbracket \pi_u(\ell_j) \rrbracket \cdot \mathsf{F}^*)k')$. et donc :

$$\bigoplus_i k_i \llbracket \pi(\ell_i) \rrbracket = k\mathsf{c}(x)^* \left((\frac{\partial}{\partial a}\mathsf{F} \cdot \mathsf{F}^*)k' \right) = \frac{\partial}{\partial a}\mathsf{E} \quad .$$

$\square$

Le théorème suivant, dont la preuve est similaire à celle du Théorème 36, permet de dériver des concaténations à droite de sous-expressions – unitaires pondérées – :

Théorème 48. *Soit* E *une expression rationnelle et* X *un mot sur les nœuds de* T_{E}*. Si* $\mathsf{First}(X, a) = \bigoplus_i k_i \ell_i$*, alors :*

$$\frac{\partial}{\partial a} \llbracket X \rrbracket = \bigoplus_i k_i \llbracket \pi(\ell_i) \rrbracket \quad .$$

Pour une feuille propre h telle que $\mathsf{First}(\pi(h), a) = \bigoplus_i k_i \ell_i$, on a donc :

$$\frac{\partial}{\partial a} \llbracket \pi(h) \rrbracket = \bigoplus_i k_i \llbracket \pi(\ell_i) \rrbracket \quad .$$

Et finalement, directement :

Théorème 49.

$$\mathrm{TD}\,(\mathsf{E}) = \{ \llbracket \pi(\ell) \rrbracket \mid \ell \in \mathrm{fp}(\mathsf{E}) \}$$

Construction de l'automate

Le théorème précédent, identique au cas booléen, permet, comme dans ce cas là, de considérer que les états $\mathcal{D}(\mathsf{E})$ ne sont plus les termes dérivés mais les mots réduits sur les marques qui les représentent. L'unique état initial est $\overline{r}$, la fonction finale est définie par le terme constant des mots de marques. Ainsi nous notons également $\theta(q)$ l'ensemble des feuilles propres ℓ de T_{E} dont la réduction du mot marqué associé au chemin est q. C'est-à-dire que si x et x' sont dans $\theta(q)$ alors : $\rho\left(\overline{\pi(x)}\right) = \rho\left(\overline{\pi(x')}\right) = q$. Un représentant x arbitrairement choisi dans $\theta(q)$ permet de définir $\mathsf{First}(q) = \mathsf{First}(\pi(x))$, justifié par :

Proposition 50. *Soit* q *un état, si* x *et* x' *sont dans* $\theta(q)$ *et si* $\mathsf{First}(\pi(x), a) = \bigoplus_i k_i \ell_i$ *et* $\mathsf{First}(\pi(x'), a) = \bigoplus_j k'_j \ell'_j$ *alors :*

$$\bigoplus_i k_i \llbracket \pi(\ell_i) \rrbracket = \bigoplus_j k'_j \llbracket \pi(\ell'_j) \rrbracket \quad .$$

Enfin nous pouvons décrire de manière constructive l'ensemble des transitions de l'automate $\mathcal{D}(\mathsf{E})$: soient q_i et q_j deux états. Si $\mathsf{First}(\pi(q_i), a) = \bigoplus_m k_m \ell_m$, alors on a par le Théorème 48 :

$$a\mu_{i,j} = \sum_{\ell_n \in \theta(q_j)} k_n \quad .$$

Cette équation permet de construire les transitions par le Théorème 46.

5.3 Termes dérivés cassés

Dans cette section, nous donnons finalement la définition des termes dérivés cassés pour le cas des expressions rationnelles à multiplicités. Une définition avait déjà été envisagé dans [31] mais ce n'est pas la définition que nous retenons et nous nous justifierons de ce choix. Nous achevons cette partie sur l'adaptation des constructions données précédemment afin de construire l'automate des termes dérivés cassés pour les expressions rationnelles à multiplicités.

5.3.1 Définition

Le cassage des expressions booléennes peut être interprété comme une dérivation par le mot vide. Il en résulte que le cassage d'une expression rationnelle booléenne donne un ensemble d'expressions rationnelles. Dans le cas des expressions à multiplicités, le cassage d'une expression sera donc, comme la dérivation, une combinaison linéaire à gauche d'expressions rationnelles.

Il nous faut alors quelques petites notations préliminaires. Soit X une combinaison linéaire à gauche d'expressions rationnelles. Le coefficient de l'expression $\mathsf{1}$ dans X est noté δ_X et la combinaison linéaire égale a X sauf pour l'expression $\mathsf{1}$ qui a un coefficient nul est notée $(X)_{\mathsf{p}}$. Il en découle entre autres que $(X \oplus Y)_{\mathsf{p}} = X_{\mathsf{p}} \oplus Y_{\mathsf{p}}$ et $\delta_{X_{\mathsf{p}}} = \mathsf{0}$ pour tout X et Y. Ces notations sont évidemment cohérentes avec celles posées dans le cas booléen.

Définition 25. *Le* cassage *de l'expression rationnelle* E *est la combinaison linéaire à gauche* $\mathrm{B}(\mathsf{E})$ *définie inductivement par :*

$$\mathrm{B}(\mathsf{0}) = \mathsf{0}\,, \quad \mathrm{B}(\mathsf{1}) = \mathsf{1}\,, \quad \forall a \in A \quad \mathrm{B}(a) = a\,,$$
$$\mathrm{B}(\mathsf{F}+\mathsf{G}) = \mathrm{B}(\mathsf{F}) \oplus \mathrm{B}(\mathsf{G}) \quad ,$$
$$\mathrm{B}(\mathsf{F}\cdot\mathsf{G}) = (\mathrm{B}(\mathsf{F}))_{\mathsf{p}} \cdot \mathsf{G} \oplus \delta_{\mathrm{B}(\mathsf{F})}\mathrm{B}(\mathsf{G}) \quad ,$$
$$\mathrm{B}(\mathsf{F}^*) = \mathsf{F}^* \quad ,$$
$$\mathrm{B}(k\mathsf{F}) = k\mathrm{B}(\mathsf{F}) \;, \quad \mathrm{B}(\mathsf{F}k) = (\mathrm{B}(\mathsf{F})\,k) \quad .$$

Le cassage des expressions s'étend naturellement au cassage de combinaisons linéaires d'expressions par :

$$\mathrm{B}\left([kX]\right) = k\mathrm{B}\left(X\right) \qquad \text{et} \qquad \mathrm{B}\left(X \oplus Y\right) = \mathrm{B}\left(X\right) \oplus \mathrm{B}\left(Y\right) \quad .$$

Exemple 52 (*Ex. 38 cont.*)**.** Le cassage de l'expression E_1 donne : $\mathrm{B}\left(\mathsf{E}_1\right) = [6(\mathsf{F}_1 6)] \oplus [3(\mathsf{F}_1^* 2)]$

Nous pouvons remarquer que nous n'avons pas appelé cette fonction termes cassés, comme c'est le cas dans le cas booléen. C'est parce que nous ne voulons pas confondre les 'termes', qui sont les états de l'automate des termes dérivés cassés et donc un ensemble avec la combinaison linéaire dont les coefficients permettront de calculer les poids des transitions de l'automate.

Pour cela, par analogie avec les séries rationnelles, nous noterons $\mathsf{supp}(X)$ le support de la combinaison linéaire X, c'est-à-dire l'ensemble des expressions dont le coefficient est non nul dans X. L'ensemble $\mathsf{supp}(\mathrm{B}\left(\mathsf{E}\right))$, qui correspond donc aux expressions qui apparaissent dans le cassage de E, est alors appelé, par analogie avec le cas booléen, l'ensemble des *termes cassés* de E. Sa définition inductive est :

$$\mathsf{supp}(\mathrm{B}\left(\mathsf{0}\right)) = \emptyset\,, \quad \mathsf{supp}(\mathrm{B}\left(\mathsf{1}\right)) = \{\mathsf{1}\}\,, \quad \forall a \in A \quad \mathsf{supp}(\mathrm{B}\left(a\right)) = \{a\}\,,$$

$$\mathsf{supp}(\mathrm{B}\left(\mathsf{F}+\mathsf{G}\right)) = \mathsf{supp}(\mathrm{B}\left(\mathsf{F}\right)) \cup \mathsf{supp}(\mathrm{B}\left(\mathsf{G}\right)) \quad ,$$

$$\mathsf{supp}(\mathrm{B}\left(\mathsf{F}\cdot\mathsf{G}\right)) = \left(\mathsf{supp}(\mathrm{B}\left(\mathsf{F}\right))\right)_{\mathsf{p}} \cdot \mathsf{G} \cup \delta_{\mathsf{supp}(\mathrm{B}(\mathsf{F}))}\mathsf{supp}(\mathrm{B}\left(\mathsf{G}\right)) \quad ,$$

$$\mathsf{supp}(\mathrm{B}\left(\mathsf{F}^*\right)) = \mathsf{F}^* \quad ,$$

$$\mathsf{supp}(\mathrm{B}\left(k\mathsf{F}\right)) = \mathsf{supp}(\mathrm{B}\left(\mathsf{F}\right))\,, \quad \mathsf{supp}(\mathrm{B}\left(\mathsf{F}k\right)) = \left(\mathsf{supp}(\mathrm{B}\left(\mathsf{F}\right))k\right) \quad .$$

De par la définition, les termes cassés de E sont nécessairement des sous-expressions unitaires à gauche de l'expression E, de plus, on peut montrer par une récurrence directe que :

$$\left|\mathrm{B}\left(\mathsf{E}\right)\right| = \left|\mathsf{E}\right| \quad .$$

Définition 26. *La* $\mathbb{K}$-dérivation cassée *d'une expression rationnelle* E *sur* A *par rapport à une lettre* a *de* A *est définie par :*

$$\frac{\partial_{\mathsf{b}}}{\partial a}\mathsf{E} = \mathrm{B}\left(\frac{\partial}{\partial a}\mathsf{E}\right) \quad .$$

La $\mathbb{K}$*-derivation cassée par rapport à un mot non vide* ua *est définie par induction sur la longueur des mots :*

$$\frac{\partial_{\mathsf{b}}}{\partial ua}\mathsf{E} = \frac{\partial_{\mathsf{b}}}{\partial a}\frac{\partial_{\mathsf{b}}}{\partial u}\mathsf{E} \quad .$$

Une induction directe permet de montrer que pour toute expression E, on a :

$$\forall a \in A \quad \frac{\partial}{\partial a} \mathrm{B}(\mathsf{E}) = \frac{\partial}{\partial a} \mathsf{E}\,, \qquad \text{et donc} \qquad \forall f \in A^+ \quad \frac{\partial_{\mathsf{b}}}{\partial f} \mathsf{E} = \mathrm{B}\left(\frac{\partial}{\partial f} \mathsf{E}\right) \quad .$$

Finalement nous prenons la définition des termes dérivés cassés suivant – comme pour les termes dérivés nous prenons la définition inductive – :

Définition 27. *L'ensemble des* vraies termes dérivés cassés *est l'ensemble* $\mathrm{TBD}(\mathsf{E})$ *défini inductivement par :*

$$\begin{aligned}
\mathrm{TBD}(0) = \mathrm{TBD}(1) = \emptyset, \qquad \mathrm{TBD}(a) &= \{1\} \quad , \\
\mathrm{TBD}(\mathsf{F} + \mathsf{G}) &= \mathrm{TBD}(\mathsf{F}) \ \cup \ \mathrm{TBD}(\mathsf{G}) \quad , \\
\mathrm{TBD}(\mathsf{F} \cdot \mathsf{G}) &= (\mathrm{TBD}(\mathsf{F}))_{\mathsf{p}} \cdot \mathsf{G} \ \cup \ \mathrm{TBD}(\mathsf{G}) \ \cup \ \delta_{\mathrm{TBD}(\mathsf{F})} \mathsf{supp}(\mathrm{B}(\mathsf{G})) \quad , \\
\mathrm{TBD}(\mathsf{F}^*) &= [\mathrm{TBD}(\mathsf{F})] \cdot \mathsf{F}^* \quad , \\
\mathrm{TBD}(k\mathsf{F}) &= \mathrm{TBD}(\mathsf{F}) \ , \quad \mathrm{TBD}(\mathsf{F}k) = (\mathrm{TBD}(\mathsf{F})\, k) \quad .
\end{aligned}$$

Définition 28. *L'ensemble des* termes dérivés cassés *de* E *est :* $\mathrm{BD}(\mathsf{E}) = \mathrm{TBD}(\mathsf{E}) \cup \mathsf{supp}(\mathrm{B}(\mathsf{E}))$.

Ces définitions ont été choisies pour que l'on ait le résultat suivant :

Proposition 51. *Les termes dérivés cassés d'une expression rationnelle sont le support du cassage des termes dérivés de cette expression.*

Démonstration. Cette proposition se prouve par induction sur l'expression rationnelle. L'induction des cas de base, de la somme, de l'étoile et des multiplications par un scalaire est directe par la définition inductive de $\mathsf{supp}(\mathrm{B}(\mathsf{E}))$. Nous allons développer le cas $\mathsf{E} = \mathsf{F} \cdot \mathsf{G}$:

$$\begin{aligned}
\mathsf{supp}(\mathrm{B}(\mathrm{TD}(\mathsf{F} \cdot \mathsf{G}))) &= \mathsf{supp}(\mathrm{B}(\mathrm{TD}(\mathsf{F}) \cdot \mathsf{G} \ \cup \ \mathrm{TD}(\mathsf{G}))) \\
&= \mathsf{supp}(\mathrm{B}(\mathrm{TD}(\mathsf{F}) \cdot \mathsf{G})) \ \cup \ \mathsf{supp}(\mathrm{B}(\mathrm{TD}(\mathsf{G}))) \quad .
\end{aligned}$$

Par induction $\mathsf{supp}(\mathrm{B}(\mathrm{TD}(\mathsf{G}))) = \mathrm{TBD}(\mathsf{G})$ et on a :

$$\mathsf{supp}(\mathrm{B}(\mathrm{TD}(\mathsf{F}) \cdot \mathsf{G})) = (\mathsf{supp}(\mathrm{B}(\mathrm{TD}(\mathsf{F}))))_{\mathsf{p}} \cdot \mathsf{G} \cup \delta_{\mathsf{supp}(\mathrm{B}(\mathrm{TD}(\mathsf{F})))} \mathsf{supp}(\mathrm{B}(\mathsf{G})) \quad .$$

Par induction $(\mathsf{supp}(\mathrm{B}(\mathrm{TD}(\mathsf{F}))))_{\mathsf{p}} = (\mathrm{TBD}(\mathsf{F}))_{\mathsf{p}}$ et donc la proposition est prouvée. ∎

Proposition 52. *Les termes dérivés cassés d'une expressions rationnelle* E *sont des concaténations à droite de sous-expressions unitaires de* E *pondérées.*

Théorème 53. *Soit* $\mathrm{BD}\,(\mathsf{E}) = \{\mathsf{K}'_1, ..., \mathsf{K}'_n\}$ *l'ensemble des termes dérivés cassés de* E. *Pour toute lettre* a *de* A^*, *il existe une matrice carrée* $a\mu'$, *de dimension* n *et à coefficients dans* $\mathbb{K}$ *telle que :*

$$\forall i \qquad \frac{\partial_{\mathsf{b}}}{\partial a}\,\mathsf{K}'_i = \bigoplus_j a\mu'_{i,j}\,\mathsf{K}'_j\ .$$

Démonstration. Il suffit, pour démontrer ce théorème, de montrer que le support de la dérivation cassante d'un terme dérivé cassé est un ensemble de termes dérivés cassés. On sait déjà, par le Théorème 46, que le support de la dérivation des termes dérivés est un ensemble de termes dérivés. Or comme les termes dérivés cassés sont le support du cassage des termes dérivés, en passant au cassage de la dérivation, on obtient bien le résultat. □

Finalement nous pouvons donner la définition naturelle de l'automate des termes dérivés cassés :

Définition 29. L'automate des termes dérivés cassés *d'une expression* E *est l'automate* $\mathcal{D}_b(\mathsf{E})$ *dont les états sont les termes dérivés cassés.*

(i) *Les états initiaux sont les termes cassés de* E *et le poids associé est celui de l'expression dans le cassage* $\mathrm{B}\,(\mathsf{E})$.

(ii) *La fonction finale est le terme constant de chaque terme dérivé cassé.*

(iii) *Il y a une transition de* K_i *à* K_j *étiqueté par* ka *si* $k = a\mu'_{i,j} \neq 0_{\mathbb{K}}$.

Cet automate, reconnaît la série dénoté par E :

Théorème 54. *Le comportement de l'automate des termes dérivés cassés* $\mathcal{D}_b(\mathsf{E})$ *de l'expression* E *est la série dénotée par* E.

Démonstration. Nous montrons que le comportement de l'automate des termes dérivés cassés est le même que celui des termes dérivés. En effet, si un mot u a pour coefficient k dans la série reconnue par l'automate des termes dérivés d'une expression rationnelle E, cela veut dire que :

$$\mathsf{c}(\frac{\partial}{\partial u}\,\mathsf{E}) = k\ \ .$$

Comme le comportement du cassage d'une expression est le comportement de l'expression on a :

$$\mathsf{c}(\frac{\partial_{\mathsf{b}}}{\partial u}\,\mathsf{E}) = k\ \ .$$

Cela signifie que la somme de tous les coefficients des calculs réussis de u dans l'automate des termes dérivés cassés de E est k et donc le comportement de l'automate est le même que celui des termes dérivé : c'est la série dénotée par E. □

Exemple 53. Prenons l'exemple de l'expression $\mathsf{E}_3 = 2((3a+b)5 \cdot 2(c+d))$. Le cassage de l'expression E_3 donne :

$$\mathrm{B}\,(\mathsf{E}_3) = [6(a5 \cdot 2(c+d))] \oplus [2(b5 \cdot 2(c+d))] \ ,$$
$$\mathsf{supp}(\mathrm{B}\,(\mathsf{E}_3)) = \{a5 \cdot 2(c+d) \ , \ b5 \cdot 2(c+d)\} \ .$$

Les vrais termes dérivés cassés de E_3 sont :

$$\mathrm{TBD}\,(\mathsf{E}_3) = \{c \ , \ d \ , \ 1\} \ .$$

La dérivation des termes cassés donne :

$$\frac{\partial_{\mathsf{b}}}{\partial a}\, a5 \cdot 2(c+d) = [10c] \oplus [10d]$$
$$\frac{\partial_{\mathsf{b}}}{\partial b}\, b5 \cdot 2(c+d) = [10c] \oplus [10d]$$

L'automate des termes dérivés cassés de E_3 est donc l'automate de la Figure 5.5.

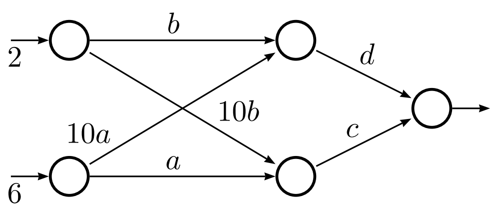

FIGURE 5.5 – L'automate des termes dérivés cassés de E_3

5.3.2 Calcul de l'automate

Une nouvelle fois, nous allons modifier la méthode de calcul des termes dérivés pour obtenir l'automate des termes dérivés cassés. Pour ce faire, comme la dérivation cassée est le cassage de la dérivation, il suffit de définir, comme dans le cas booléen, une opération de cassage des mots sur les u-nœuds pondérés.

Cassage de mots d'u-nœuds pondérés

Afin de définir le cassage de mots, nous introduisons d'abord la combinaison linéaire d'u-nœuds pondérés $\omega(x)$, où $x = (k, u, k')$, par :

$$\omega(x) = \begin{cases} k(1_{\mathbb{K}}, u, k') & \text{si } u \text{ est une feuille ou si } u = y^*, \\ k\omega(y) & \text{si } u = y \cdot z, \\ k\omega(y) \oplus k\omega(z) & \text{si } u = y + z. \end{cases}$$

Le *cassage* du mot X sur les u-nœuds pondérés est défini inductivement par la combinaison linéaire à gauche de mots suivante :

(i) $\mathrm{B}(X) = \varepsilon$ si $X = \varepsilon$

(ii) si $X = x_1 :: \ldots :: x_n$ où $x_i = (k_i, u_i, k'_i)$ et si $\omega(x_1) = \bigoplus_j t_j y_j$ où $y_j = (h_j, v_j, h'_j)$ alors :

$$\begin{aligned} \mathrm{B}(X) = & \bigoplus_{v_j \neq 1} t_j (y_j :: \pi_{x_1}(y_j) :: x_2 :: \ldots :: x_n) \\ & \bigoplus_{v_j = 1} t_j \mathrm{B}(\pi_{x_1}(y) :: x_2 :: \ldots :: x_n) \ . \end{aligned} \tag{5.5}$$

Cette définition de cassage des mots se justifie par le théorème suivant :

Théorème 55. *Soit X un mot sur les u-nœuds pondérés de l'arbre T_{E}. Si $\mathrm{B}(X) = \bigoplus_i k_i Y_i$, alors :*

$$\mathrm{B}(\llbracket X \rrbracket) = \bigoplus_i k_i \llbracket Y_i \rrbracket \ .$$

Démonstration. Le raisonnement de cette preuve est le même que celui dans le cas booléen. Pour simplifier l'écriture, nous allons uniquement traiter les cas qui ne sont pas traités dans le cas booléen, c'est-à-dire :

Si $\llbracket X \rrbracket = (k\mathsf{E})$, alors on a nécessairement X réduit à un seul u-nœud pondéré de la forme (k, u, k'). Par définition de $\omega(u)$, et comme le chemin d'un nœud ne récupère pas les coefficients à gauche, on a $\mathrm{B}((k, u, k')) = k\mathrm{B}((1_{\mathbb{K}}, u, k'))$. Comme $\llbracket (1_{\mathbb{K}}, u, k') \rrbracket = \mathsf{E}$ alors l'énoncé est vérifié.

Si $\llbracket X \rrbracket = (\mathsf{E}k)$, alors on a $X = x_1 :: \ldots :: x_n :: k$ ou bien $X = (1_{\mathbb{K}}, u, k)$. Dans le premier cas, la définition récursive du cassage de mot assure que $\mathrm{B}(X)) = \mathrm{B}(x_1 :: \ldots :: x_n) :: k$. Dans le deuxième cas, c'est la définition du chemin qui récupère les coefficients à droite qui assure que $\mathrm{B}(X)) = \mathrm{B}((1_{\mathbb{K}}, u, 1_{\mathbb{K}})) :: k$. Dans ces deux cas l'énoncé se vérifie.

□

En particulier, si r est la racine de T_{E}, nous avons $\mathrm{B}(\mathsf{E}) = \mathrm{B}(r)$ et l'ensemble des termes cassés de E sont le support de $\mathrm{B}(r)$.

L'application du Théorème 55 et des résultats pour les termes dérivés cassés donne directement :

Théorème 56. *Soit E une expression rationnelle et X un mot sur les u-nœuds pondérés. Si $\mathsf{First}(X, a) = \bigoplus_i k_i \ell_i$ et $\mathrm{B}(\pi(\ell_i)) = \bigoplus_j h_{i,j} X_{i,j}$, alors :*

$$\frac{\partial_{\mathsf{b}}}{\partial a} \llbracket X \rrbracket = \bigoplus_{i,j} k_i h_{i,j} \llbracket X_{i,j} \rrbracket \ .$$

Théorème 57.

$$\mathrm{TD}\,(\mathsf{E}) = \bigcup_{\ell \in \mathrm{fp}(\mathsf{E})} \mathsf{supp}(\{[\![X]\!] \mid X \in \mathrm{B}(\pi(l))\})$$

Construction de l'automate

Il va donc être possible de construire l'automate des termes dérivés cassés de l'expression rationnelle par la représentation sous forme de mots sur les u-nœuds pondérés. Pour cela :

1. calcul pour tout nœud $x = (k, u, k')$ de l'arbre syntaxique – par un parcours de bas en haut – :
   – du terme constant $\mathsf{c}(x)$ et du terme constant $\mathsf{c}(u)$;
   – de l'ensemble des feuilles propres $\mathsf{First}(u)$;
   – de la marque associée $\overline{u}$;
2. calcul pour tout nœud x de l'arbre – par un parcours de haut en bas – de l'ensemble $\omega(x)$;
3. calcul du chemin $\pi(\ell)$ associé à toute feuille propre ℓ deT_{E} ; calcul du cassage $\mathrm{B}(\pi(\ell))$.
4. Pour tout mot de $\mathsf{supp}(\mathrm{B}(\pi(\ell)))$: calcul du mot réduit par ρ – par calculs successifs par ρ_1, ρ_2, ρ_3, ρ_4 et ρ_3 –. Pour chaque mot X il faut retenir l'ensemble des feuilles ℓ telles que $X \in \mathsf{supp}(\mathrm{B}(\pi(\ell)))$ ainsi que son coefficient dans le mot $\mathrm{B}(\pi(\ell))$;
5. identification des mots X et X' obtenus à l'étape précédente, tels que $\rho\left(\overline{X}\right) = \rho\left(\overline{X'}\right)$. Cette étape donne les états de l'automate, il faut garder la composition de chaque classe $\theta_b(q)$ de mots partageant la même marque ;
6. pour chaque état q, calcul de $\mathsf{First}(X)$ pour un représentant $X \in \theta_b(q)$ donné ;
7. identification des états finaux et de leur poids par le calcul de $\mathsf{c}(q)$ pour tous les états q ;
8. la fonction initiale est donnée par $\mathrm{B}(r)$;
9. il existe une transition (q_i, ka, q_j) si les trois conditions suivantes sont satisfaites :
   – il existe une feuille ℓ étiquetée par a avec un coefficient k' non nul dans $\mathsf{First}(q_i, a)$,
   – il existe $X \in \theta_b(q_j)$ avec un coefficient k'' non nul dans $\mathrm{B}(\pi(l))$ tel que $k'k'' = k$.