DM de Logique #1, 2019

Preuves Infinies pour les Algèbres de Kleene

David Baelde, ENS Paris-Saclay

Ce devoir est obligatoire, à rédiger individuellement pour le 3 avril.
Les questions marquées (?) nécessitent plus d'intuition ;
celles marquées (!) sont un peu plus techniques.

On considère les expressions régulières sur un alphabet A, données par la syntaxe suivante :

$$e, f ::= e \cdot f \mid e + f \mid e^* \mid 1 \mid 0 \mid a \in A$$

Chaque expression dénote un langage $\mathcal{L}(e) \subseteq A^*$.

On peut aussi interpréter les expressions dans des structures plus générales, appelées algèbres de Kleene [1], et se demander si une égalité $e = f$ est valide, i.e. si e et f ont les mêmes interprétations $[\![e]\!]_\mathcal{A}$ et $[\![f]\!]_\mathcal{A}$ dans toute algèbre de Kleene. En fait, il s'avèrera plus naturel de travailler sur les inégalités, en définissant $e \leq f$ comme $e + f = f$, ce qui signifie $\mathcal{L}(e) \subseteq \mathcal{L}(f)$ quand on considère l'interprétation dans l'algèbre des langages sur l'alphabet A. De façon remarquable, il a été montré que les inégalités vraies dans l'algèbre des langages sont exactement celles vraies dans toutes les algèbres de Kleene, ce qui permet de ramener le problème de la validité au problème de l'inclusion de langages. Ce résultat nous fournit donc une procédure de décision pour la validité, par exemple via les automates. On peut aussi chercher à appréhender ce problème via la théorie de la preuve ; c'est ce qui nous intéresse ici.

Dans ce contexte, on considère des séquents de la forme $\Gamma \vdash e$ où e est une expression régulière et Γ est une *liste* d'expressions régulières. La notation Γ, Δ dénote la concaténation de listes. Si $\Gamma = (e_1, \ldots, e_n)$ on notera $\Pi\Gamma := e_1 \cdot \ldots \cdot e_n$ et on dira qu'un séquent $\Gamma \vdash e$ est *valide* quand $\mathcal{L}(\Pi\Gamma) \subseteq \mathcal{L}(e)$. Dans la suite, on écrira simplement $\mathcal{L}(\Gamma)$ pour $\mathcal{L}(\Pi\Gamma)$.

Question 1

Parmi les séquents suivants, lesquels sont valides ?

1. $a + b, a + b \vdash a \cdot b$
2. $a, b \vdash (a + b) \cdot (a + b)$
3. $a, (a + b)^* \cdot b, (b \cdot a)^* \vdash b^* \cdot (a \cdot b^*)^*$
4. $b^* \cdot (a \cdot b^*)^* \vdash a \cdot (a + b)^* \cdot b \cdot (b \cdot a)^*$

1. https://en.wikipedia.org/wiki/Kleene_algebra

Une première approche pour construire un calcul des séquents pour les algèbres de Kleene est d'inclure une règle permettant de raisonner par induction sur l'étoile. C'est possible, mais il est difficile d'obtenir un calcul sans coupure. Dans tous les cas, la recherche de preuve est difficile dans les calculs obtenus, puisqu'elle va nécessiter l'invention d'invariants potentiellement complexes. Montrer la complétude de tels calculs est aussi notoirement difficile.

Une autre approche consiste à n'utiliser que des règles de raisonnement locales, en éliminant le raisonnement par induction sur l'étoile, mais en permettant aux dérivations d'être infinies : on passe d'arbres finis à des arbres à branchement borné mais profondeur potentiellement infinie. De telles preuves constituent un outil théorique intéressant, facilitant typiquement les preuves de complétude. De plus, si l'on arrive à montrer que les preuves régulières [2] prouvent les mêmes séquents que les preuves infinies, alors on a même un outil pratique utilisable par exemple pour faire de la recherche de preuve.

$$
\frac{}{e \vdash e}\,\mathsf{ax} \qquad \frac{}{\Gamma, 0, \Delta \vdash e}\,0_{\mathsf{L}} \qquad \frac{\Gamma, \Delta \vdash e}{\Gamma, 1, \Delta \vdash e}\,1_{\mathsf{L}} \qquad \frac{}{\vdash 1}\,1_{\mathsf{R}}
$$

$$
\frac{\Gamma, e_1, \Delta \vdash f \quad \Gamma, e_2, \Delta \vdash f}{\Gamma, e_1 + e_2, \Delta \vdash f}\,+_{\mathsf{L}} \qquad \frac{\Gamma \vdash e_i}{\Gamma \vdash e_1 + e_2}\,+_{\mathsf{R}i}
$$

$$
\frac{\Gamma, e_1, e_2, \Delta \vdash f}{\Gamma, e_1 \cdot e_2, \Delta \vdash f}\,\cdot_{\mathsf{L}} \qquad \frac{\Gamma \vdash e \quad \Delta \vdash f}{\Gamma, \Delta \vdash e \cdot f}\,\cdot_{\mathsf{R}}
$$

$$
\frac{\Gamma, \Delta \vdash f \quad \Gamma, e, e^*, \Delta \vdash f}{\Gamma, e^*, \Delta \vdash f}\,*_{\mathsf{L}} \qquad \frac{}{\vdash e^*}\,*_{\mathsf{R1}} \qquad \frac{\Gamma \vdash e \cdot e^*}{\Gamma \vdash e^*}\,*_{\mathsf{R2}}
$$

Figure 1 – Règles d'inférence pour LKA

Nous aller étudier un tel système, appelé LKA. Une *pré-preuve* de LKA est une dérivation potentiellement infinie utilisant les règles de la figure 1. Plus formellement, c'est un arbre potentiellement infini, dont chaque noeud n est étiqueté par un séquent et le nom d'une règle d'inférence, de sorte que les séquents du noeud n et de ses noeuds fils constituent une instance de la règle d'inférence. Les pré-preuves ne sont pas correctes par rapport à notre sémantique. Par exemple, voici une pré-preuve (infinie et régulière) du séquent invalide $a \vdash 1^*$:

$$
\dfrac{\dfrac{}{\vdash 1}\,1_{\mathsf{R}} \qquad \dfrac{\dfrac{}{\vdash 1}\,1_{\mathsf{R}} \quad \dfrac{\vdots}{a \vdash 1^*}}{a \vdash 1^*}\,*_{\mathsf{R2}}}{a \vdash 1^*}\,*_{\mathsf{R2}}
$$

On définit enfin les *preuves* comme les pré-preuves dans lesquelles toute branche infinie

2. Un arbre est régulier s'il n'a qu'un nombre fini de sous-arbres, i.e. s'il peut être représenté par un graphe orienté connexe.

contient une infinité d'occurrences de la règle $*_\mathsf{L}$. La pré-preuve ci-dessus n'est pas une preuve, et nous allons montrer que les preuves sont correctes.

Question 2
Donner une preuve du séquent $(a \cdot b + b \cdot a)^* \vdash b^* \cdot (a \cdot b^*)^*$.

1 Correction

L'objectif de cette première partie est de montrer que les preuves de LKA sont correctes. Comme les dérivations sont infinies, la correction locale des règles ne suffit pas à conclure : il nous faudra donc un argument supplémentaire.

Question 3
Montrer que les règles $*_\mathsf{L}$, $*_\mathsf{R1}$ et $*_\mathsf{R2}$ sont correctes. On admettra pour la suite la correction de toutes les règles.

Question 4 (?!)
On pose $e^0 = 1$, $e^{n+1} = e \cdot (e^n)$.
Montrer que si $\Gamma, e^*, \Delta \vdash_\mathsf{LKA} f$ alors $\Gamma, e^n, \Delta \vdash_\mathsf{LKA} f$ pour tout $n \in \mathbb{N}$.

Question 5
Montrer que $\Gamma, e^*, \Delta \vdash f$ est valide si, pour tout n, le séquent $\Gamma, e^n, \Delta \vdash f$ est valide.

Étant donné une expression e on définit sa hauteur d'étoile $h(e)$ et sa multiplicité $m(e)$ comme suit :

$$\begin{array}{rcl} h(0) & = & h(1) = h(a) = 0 \\ h(e \cdot f) & = & h(e+f) = \max(h(e), h(f)) \\ h(e^*) & = & 1 + h(e) \end{array}$$

$$\begin{array}{rcl} m(0) & = & m(1) = m(a) = m(e^*) = 1 \\ m(e \cdot f) & = & m(e+f) = m(e) + m(f) \end{array}$$

On définit enfin le poids d'une liste Γ comme le multi-ensemble des hauteurs de ses éléments, avec les multiplicités définies plus haut. Formellement, en représentant les multi-ensembles comme des fonctions de $\mathbb{N} \to \mathbb{N}$:

$$w(e_1, \ldots, e_n) = k \mapsto \Sigma_{i \in [1;n], h(e_i) = k} m(e_i)$$

Par exemple, $w(a^* \cdot b^*, a^*, b^{**})$ est la fonction f tel que $f(1) = 3$, $f(2) = 1$ et $f(n) = 0$ sinon. Les multi-ensembles seront ordonnés par l'extension multi-ensembliste de l'ordre sur les entiers naturels.

Question 6
Montrer que, si $\Gamma \vdash e$ est conclusion d'une inférence autre que $*_\mathsf{L}$ avec $\Delta \vdash f$ parmi les prémisses, alors $w(\Delta) \leq w(\Gamma)$.

Question 7
Montrer que $w(\Gamma, e^n, \Delta) < w(\Gamma, e^*, \Delta)$.

Question 8
Soit Π une dérivation de LKA. Montrer qu'il existe un préfixe [3] fini [4] de Π ne contenant aucune application de $*_\mathsf{L}$ et dont les feuilles sont soit justifiées par une règle initiale soit conclusion de $*_\mathsf{L}$ dans Π.

Question 9 (?)
En utilisant la construction précédente, et en raisonnant par induction sur $w(\Gamma)$, montrer que si $\Gamma \vdash_{\text{LKA}} e$ alors $\Gamma \vdash e$ est valide.

2 Complétude

On s'intéresse maintenant à la complétude du calcul. Nous allons montrer la variante la plus faible de ce résultat, établissant que tout séquent valide a une preuve, sans aucune information sur la forme de cette preuve. Il serait plus intéressant de pouvoir montrer que tout séquent valide admet une preuve régulière, mais c'est faux pour ce calcul simple. D'autres calculs permettent d'obtenir cela. Un autre permet même d'obtenir à partir d'une preuve régulière une preuve finie utilisant les axiomes des algèbres de Kleene. Mais ces résultats récents sont nettement plus complexes.

Question 10 (?!)
Montrer que, si on applique autant que possible (éventuellement à l'infini) des règles gauches sur un séquent $\Gamma \vdash e$, on obtient une dérivation dont les séquents non justifiés sont de la forme $a_1, \ldots, a_n \vdash e$ avec $a_1 \ldots a_n \in \mathcal{L}(\Gamma)$. Montrer de plus que chaque branche infinie de cette dérivation contient une infinité d'applications de $*_\mathsf{L}$.

Question 11 (?)
Conclure que LKA est complet : tout séquent valide est prouvable.

Question 12
Montrer que le séquent $a, a^* \vdash a^* \cdot a$ n'admet pas de preuve régulière.

3. Si un arbre est représenté via un ensemble de positions $T \subseteq \mathbb{N}^*$ tel que pour tout $u \in \mathbb{N}^*$ et $n \in \mathbb{N}$ tel que $un \in T$, on a aussi $u \in T$ et $um \in T$ pour tout $m \leq n$, alors un préfixe de T est simplement un sous-ensemble de T satisfaisant la même condition.
4. On demande simplement que le préfixe soit un arbre fini.