Décidabilité de la hiérarchie booléenne sur $\Sigma_1[< +MOD]$

Luc Dartois
sous la direction de Jean-Eric Pin et Olivier Carton

LIAFA

Hiérarchie Booléenne

- Soit S un ensemble, et $\mathcal{F}$ un treillis sur S.
  Pour $n \geqslant 1$, on appelle $\mathcal{F}(n)$ l'ensemble des sous-ensembles de S qui peuvent s'écrire

$$L = L_1 - (L_2 - (\cdots - L_n))$$

où $L_1, \cdots, L_n \in \mathcal{F}$ et $L_1 \supseteq L_2 \supseteq \cdots \supseteq L_n$.

Hiérarchie Booléenne

- Soit S un ensemble, et $\mathcal{F}$ un treillis sur S.
  Pour $n \geqslant 1$, on appelle $\mathcal{F}(n)$ l'ensemble des sous-ensembles de S qui peuvent s'écrire

$$L = L_1 - (L_2 - (\cdots - L_n))$$

  où $L_1, \cdots, L_n \in \mathcal{F}$ et $L_1 \supseteq L_2 \supseteq \cdots \supseteq L_n$.
- $\mathcal{F}(n) \subseteq \mathcal{F}(n+1)$ pour tout n
- On a $\bigcup_n \mathcal{F}(n) = \mathcal{BF}$.

Hiérarchie Booléenne

- Soit S un ensemble, et $\mathcal{F}$ un treillis sur S.
  Pour $n \geqslant 1$, on appelle $\mathcal{F}(n)$ l'ensemble des sous-ensembles de S qui peuvent s'écrire

$$L = L_1 - (L_2 - (\cdots - L_n))$$

  où $L_1, \cdots, L_n \in \mathcal{F}$ et $L_1 \supseteq L_2 \supseteq \cdots \supseteq L_n$.
- $\mathcal{F}(n) \subseteq \mathcal{F}(n+1)$ pour tout n
- On a $\bigcup_n \mathcal{F}(n) = \mathcal{BF}$.
- La question que l'on se pose est étant donné un langage L dans $\mathcal{BF}$, peut-il être décomposé de cette façon en n langages de notre logique ?

Chaînes alternées

Soit un ensemble muni d'un ordre partiel $(S, \leqslant)$, et soit $\mathcal{F}$ le treillis des ensembles clos par le haut selon l'ordre $\leqslant$.

- **Définition**

  Une chaîne L-alternée de longueur n, pour $L \subseteq S$, est une suite $(x_0, \cdots, x_n)$ d'éléments de S tels que $x_0 \leqslant \cdots \leqslant x_n$ et $x_i \in L \Leftrightarrow x_{i+1} \notin L$.

  Elle sera dite positive si $x_0 \in L$ et négative sinon.

Chaînes alternées

Soit un ensemble muni d'un ordre partiel $(S, \leqslant)$, et soit $\mathcal{F}$ le treillis des ensembles clos par le haut selon l'ordre $\leqslant$.

- **Définition**

  Une chaîne L-alternée de longueur n, pour $L \subseteq S$, est une suite $(x_0, \cdots, x_n)$ d'éléments de S tels que $x_0 \leqslant \cdots \leqslant x_n$ et $x_i \in L \Leftrightarrow x_{i+1} \notin L$.
  Elle sera dite positive si $x_0 \in L$ et négative sinon.

- **Proposition**

  Pour tout $L \subseteq S$ et $n \leqslant 1$, on a
  $L \in \mathcal{F}(n)$ si, et seulement si il n'existe pas de chaîne L-alternée positive de longueur n.

La logique $\Sigma_1[< +MOD]$

- ## Le fragment Σ_1

  Il s'agit ici d'un fragment de la logique du premier ordre sur les langages de mots.
  Les formules de Σ_1 sont composées de la façon suivante :

$$\exists x_1 \cdots \exists x_n \varphi \text{ où } \varphi \text{ est sans quantificateurs}$$

La logique $\Sigma_1[< +MOD]$

- ## Le fragment Σ_1

Il s'agit ici d'un fragment de la logique du premier ordre sur les langages de mots.
Les formules de Σ_1 sont composées de la façon suivante :

$$\exists x_1 \cdots \exists x_n \varphi \text{ où } \varphi \text{ est sans quantificateurs}$$

- On considère ici la signature suivante

Signature [<+MOD]

1. Le prédicat binaire $<$, dénotant l'ordre usuel sur les mots.
2. Les prédicats unaires a, pour chaque lettre a de l'alphabet A.
3. Les prédicats unaires MOD_r^d, avec $d > r$, vrai aux positions égales à r modulo d.

Propriétés basiques de $\Sigma_1[< +MOD]$

- Soit L un langage de A^*, on a alors équivalence des propriétés suivantes :
  1. $L \in \Sigma_1[< +MOD]$
  2. L est union finie de langages de la forme $(A^d)^* a_1 (A^d)^* \cdots a_k (A^d)^*$

Propriétés basiques de $\Sigma_1[< +MOD]$

- Soit L un langage de A^*, on a alors équivalence des propriétés suivantes :
  1. $L \in \Sigma_1[< +MOD]$
  2. L est union finie de langages de la forme $(A^d)^* a_1 (A^d)^* \cdots a_k (A^d)^*$

- Définition de $\leqslant^d$

  On définit l'ordre $\leqslant^d$ sur les mots comme l'ordre sous-mot modulo d.

  $u \leqslant^d v$ si, et seulement si, il existe une décomposition $v = v_0 u_1 v_1 \dots u_k v_k$ telle que $u = u_1 \dots u_k$ et pour tout $i \leqslant k$, $\mid v_i \mid \equiv 0 \pmod d$.

  Les langages clos par le haut selon l'ordre $\leqslant^d$ sont exactement les langages définissables par $\Sigma_1[< +MOD_d]$.

Exemples

$L = (ab)^*$

On a $(ab)^* \in \Sigma_1[< +MOD](2)$.
En effet,
$(ab)^* = (A^2)^* \setminus \left((A^2)^* b (A^2)^* A (A^2)^* \cup (A^2)^* A (A^2)^* a (A^2)^*\right)$.

Le langage a^*

De la même façon, $a^* = A^* \setminus A^* b A^*$.
On a $a^* \in \Sigma_1[< +MOD](2)$.

Indice de Stabilité

On appelle timbre un morphisme d'un monoïde généré de façon finie sur un monoïde fini.

Définition

Soit $\varphi : A^* \to M$ un timbre et $S = \varphi(A)$.
L'indice de stabilité d'un timbre est le plus petit entier s tel que $\varphi(A^s) = \varphi(A^{2s})$.

l'ensemble $\varphi(A^s)$ est appelé le semigroupe stable de φ.

Indice de Stabilité

On appelle timbre un morphisme d'un monoïde généré de façon finie sur un monoïde fini.

Définition

Soit $\varphi : A^* \to M$ un timbre et $S = \varphi(A)$.
L'indice de stabilité d'un timbre est le plus petit entier s tel que $\varphi(A^s) = \varphi(A^{2s})$.

l'ensemble $\varphi(A^s)$ est appelé le semigroupe stable de φ.

- ## Indice de stabilité d'un langage

  Soit $L \subseteq A^*$ un langage rationnel, on appelle indice de stabilité de L l'indice de stabilité minimal sur l'ensemble des morphismes reconnaissant L.

Indice de Stabilité

On appelle timbre un morphisme d'un monoïde généré de façon finie sur un monoïde fini.

Définition

Soit $\varphi : A^* \to M$ un timbre et $S = \varphi(A)$.
L'indice de stabilité d'un timbre est le plus petit entier s tel que $\varphi(A^s) = \varphi(A^{2s})$.

l'ensemble $\varphi(A^s)$ est appelé le semigroupe stable de φ.

- ## Indice de stabilité d'un langage

  Soit $L \subseteq A^*$ un langage rationnel, on appelle indice de stabilité de L l'indice de stabilité minimal sur l'ensemble des morphismes reconnaissant L.

  - Pour un langage rationnel, l'indice de stabilité est égal à celui de son timbre syntaxique.

$\Sigma_1[< +MOD]$ et indice de stabilité

- Soit $L \subseteq A^*$ un langage rationnel, et soit s son indice de stabilité. On a les propriétés suivantes :

Propriété

$L \in \Sigma_1[< +MOD] \Leftrightarrow L \in \Sigma_1[< +MOD_s]$
$L \in \mathcal{B}\Sigma_1[< +MOD] \Leftrightarrow L \in \mathcal{B}\Sigma_1[< +MOD_s]$

$\Sigma_1[< +MOD]$ et indice de stabilité

- Soit $L \subseteq A^*$ un langage rationnel, et soit s son indice de stabilité. On a les propriétés suivantes :

Propriété

$L \in \Sigma_1[< +MOD] \Leftrightarrow L \in \Sigma_1[< +MOD_s]$
$L \in \mathcal{B}\Sigma_1[< +MOD] \Leftrightarrow L \in \mathcal{B}\Sigma_1[< +MOD_s]$

- Preuve :
  $L \in \Sigma_1[< +MOD] \Rightarrow \exists d. L \in \Sigma_1[< +MOD_d]$
  Pour tout $v = v_1 v_2 \in L$, on a $v_1 (A^d)^* v_2 \subseteq L$. Donc $v_1 (A^{sd})^* v_2 \subseteq L$. Or pour tout $u \in A^s$, il existe $w \in A^{sd}$ tel que $u \sim_L w$.
  L est clos par le haut selon l'ordre $\leqslant^s$.

Caractérisation algébrique de $\Sigma_1[< +MOD]$

Théorème [CPS06]

Soit $L \subseteq A^*$ un langage rationnel, φ son timbre syntaxique et s son indice de stabilité. Alors les propositions suivantes sont équivalentes :

1. $L \in \Sigma_1[< +MOD]$
2. $\varphi \in \mathbb{Q}\mathbf{J}^+ = [\![x \leqslant 1, x \in \varphi(A^s)]\!]$

Remarque : L'équation définissant $\mathbb{Q}\mathbf{J}^+$ redémontre directement le résultat précédent.

Propriété algébrique de $\mathcal{B}\Sigma_1[< +MOD]$

Théorème

Soit L un langage rationnel. Alors les propositions suivantes sont équivalentes :

1. $L \in \mathcal{B}\Sigma_1[< +MOD]$
2. Il existe un monoïde M reconnaissant L et un ordre $\leqslant$ sur M tels que $(M, \leqslant)$ vérifie $x \leqslant 1, x \in \varphi(A^s)$.

Preuve

- Soit $L \in \mathcal{B}\Sigma_1[< +MOD]$, il existe alors $n \leqslant 1$ et des langages $(L_i)_{i \leqslant n} \in \Sigma_1[< +MOD_s]$ tels que $L = L_1 - (L_2 - (\cdots - L_n))$. Soient $\mu_i : A^* \to M_i$ les timbres syntaxiques des langages L_i. Alors $\mu : A^* \to M = \prod_{i=1}^{n} M_i$ reconnait L.

Preuve

- Soit $L \in \mathcal{B}\Sigma_1[< +MOD]$, il existe alors $n \leqslant 1$ et des langages $(L_i)_{i \leqslant n} \in \Sigma_1[< +MOD_s]$ tels que $L = L_1 - (L_2 - (\cdots - L_n))$.
  Soient $\mu_i : A^* \to M_i$ les timbres syntaxiques des langages L_i.
  Alors $\mu : A^* \to M = \prod_{i=1}^{n} M_i$ reconnait L.
- On munit M de l'ordre produit $\leqslant$, et on note $n = ppcm_i(s_i)$.
  Alors pour tout i, $\mu_i((A^n)^*)$ vérifie l'équation $x \leqslant 1$, donc $\mu((A^n)^*)$ vérifie $x \leqslant 1$.
  Enfin, pour t indice de stabilité de μ, on a $\mu(A^t) = \mu(A^{nt}) \subseteq \mu((A^n)^*)$. $\square$

Réciproque

- Soit $\mu : A^* \to M$ d'indice de stabilité s et reconnaissant L et soit l'ordre $\leqslant$ tel que $(M, \leqslant)$ vérifie $x \leqslant 1, x \in \mu(A^s)$.
  On utilise la propriété suivante : soit $N \subseteq M$ un ensemble clos par le bas, alors $\mu^{-1}(N) \in \Sigma_1[< +MOD_s]$.

Réciproque

- Soit $\mu : A^* \to M$ d'indice de stabilité s et reconnaissant L et soit l'ordre $\leqslant$ tel que $(M, \leqslant)$ vérifie $x \leqslant 1, x \in \mu(A^s)$.
  On utilise la propriété suivante : soit $N \subseteq M$ un ensemble clos par le bas, alors $\mu^{-1}(N) \in \Sigma_1[< +MOD_s]$.
- On note alors $N_k = \{x \in M \mid$ il existe une chaîne $\mu(L)$-alternée descendante positive de longueur au moins k et finissant par $x\}$.
  Chaque N_k est clos par le bas selon l'ordre $\leqslant$ et $N_j \subseteq N_i$ pour $i \leqslant j$.

Réciproque

- Soit $\mu : A^* \to M$ d'indice de stabilité s et reconnaissant L et soit l'ordre $\leqslant$ tel que $(M, \leqslant)$ vérifie $x \leqslant 1, x \in \mu(A^s)$.
  On utilise la propriété suivante : soit $N \subseteq M$ un ensemble clos par le bas, alors $\mu^{-1}(N) \in \Sigma_1[< +MOD_s]$.
- On note alors $N_k = \{x \in M \mid$ il existe une chaîne $\mu(L)$-alternée descendante positive de longueur au moins k et finissant par $x\}$.
  Chaque N_k est clos par le bas selon l'ordre $\leqslant$ et $N_j \subseteq N_i$ pour $i \leqslant j$.
- Au final, on a
  $L = \mu^{-1}(N_0) - \mu^{-1}(N_1) - \cdots - \mu^{-1}(N_{|M|}) \in \mathcal{B}\Sigma_1[< +MOD]$.

Monoïde issu

- **Définition**
  Soit $L \in \mathcal{B}\Sigma_1[< +MOD]$. On appelle monoïde issu de L dans $\Sigma_1[< +MOD_s]$ le monoïde $M \subseteq \prod_{i=1}^{n} M_i$ où M_i est le monoïde syntaxique de L_i, i-ème composante de la décomposition de L dans $\Sigma_1[< +MOD_s]$.
  On équipe M d'un ordre $\leqslant$, clôture transitive et reflexive de la relation $x \leqslant 1$, pour $x \in S$.

Monoïde issu

- **Définition**
  Soit $L \in \mathcal{B}\Sigma_1[< +MOD]$. On appelle monoïde issu de L dans $\Sigma_1[< +MOD_s]$ le monoïde $M \subseteq \prod_{i=1}^{n} M_i$ où M_i est le monoïde syntaxique de L_i, i-ème composante de la décomposition de L dans $\Sigma_1[< +MOD_s]$.
  On équipe M d'un ordre $\leqslant$, clôture transitive et reflexive de la relation $x \leqslant 1$, pour $x \in S$.
  - On note s l'indice de stabilité de L, t celui de μ et P l'image de L par μ.
    On va maintenant prouver que la complexité minimale de L dans la hiérarchie booléenne sur $\Sigma_1[< +MOD]$ est atteinte dans $\Sigma_1[< +MOD_t]$.

Propriétés sur les chaînes

- **Propriété**

  Pour toute chaîne L-alternée selon l'ordre $\leqslant^t$, il existe une chaîne P-alternée descendante de même longueur et de même signe.

Propriétés sur les chaînes

- **Propriété**

  Pour toute chaîne L-alternée selon l'ordre $\leqslant^t$, il existe une chaîne P-alternée descendante de même longueur et de même signe.

  - Preuve :
    Soit $u_0 \leqslant^t u_1 \leqslant^t \cdots \leqslant^t u_n$ une chaîne L-alternée. En notant $x_i = \mu(u_i)$, on a $x_i \in P$ si, et seulement si $u_i \in L$.
    Enfin, pour $i \leqslant j$, on a $u_j = v_0 w_1 v_1 \dots w_k v_k$, où $v_l \in (A^t)^*$ et $u_i = w_1 \dots w_k$. Alors
    $x_j = \mu(u_j) = \mu(v_0)\mu(w_1)\mu(v_1)\dots\mu(w_k)\mu(v_k)$ où $\mu(v_l)$ est dans le semigroupe stable de M, donc $\mu(v_k) \leqslant 1$.
    On obtient donc $x_j \leqslant x_i$ pour tout $i \leqslant j$.

Propriétés sur les chaînes (2)

Propriété

Soit M un monoïde ordonné et $P \subseteq M$.
Pour tout $x_0, \ldots, x_{n+1} \in M$, si $x_0 \geqslant x_1 \geqslant \cdots \geqslant x_{n+1}$ est une chaîne P-alternée de longueur $n+1$, alors
$x_1 \geqslant \cdots \geqslant x_{n+1}$ est une chaîne $(\downarrow P - P)$-alternée de longueur n et de même signe.

Proposition

On note $L' \sqcup_d A^*$ la clôture vers le haut de L' par l'ordre $\leqslant_d$.

Proposition

Soit $L' \subseteq A^*$ tel que $\mu(L') = T$.
On a, pour tout $k > 0$, $\mu(L' \sqcup_{kt} A^*) = \downarrow T$.

Théorème clé

Théorème

Soit $T \subseteq M$ et $x_0 \geqslant x_1 \geqslant \cdots \geqslant x_n$ une chaîne T-alternée positive. Pour tout $k > 0$ et tout L' tel que $\mu(L') = T$, il existe une chaîne L'-alternée au seuil kt positive et de même longueur.

Preuve

- Prouvons le théorème par récurrence.
  Soit $x \in T$, alors x a un antécédent dans L',par définition.

Preuve

- Prouvons le théorème par récurrence.
  Soit $x \in T$, alors x a un antécédent dans L',par définition.
- Soit $x_0 \geqslant x_1 \geqslant \cdots \geqslant x_{n+1}$ une chaîne T-alternée positive de longueur $n+1$, et L' tel que $\mu(L') = T$.
  Alors $x_1 \geqslant \cdots \geqslant x_{n+1}$ est une chaîne $(\downarrow T - T)$-alternée positive de longueur n. En remarquant que $\mu(L' \sqcup_{kt} A^* - L') = \downarrow T - T$, on obtient $(u_1, \ldots, u_{n+1})$ une chaîne $(L' \sqcup_{kt} A^* - L')$-alternée positive de seuil kt.

Preuve

- Prouvons le théorème par récurrence.
  Soit $x \in T$, alors x a un antécédent dans L',par définition.
- Soit $x_0 \geqslant x_1 \geqslant \cdots \geqslant x_{n+1}$ une chaîne T-alternée positive de longueur $n+1$, et L' tel que $\mu(L') = T$.
  Alors $x_1 \geqslant \cdots \geqslant x_{n+1}$ est une chaîne $(\downarrow T - T)$-alternée positive de longueur n. En remarquant que $\mu(L' \sqcup\!\sqcup_{kt} A^* - L') = \downarrow T - T$, on obtient $(u_1, \ldots, u_{n+1})$ une chaîne $(L' \sqcup\!\sqcup_{kt} A^* - L')$-alternée positive de seuil kt.
- Or, $u_1 \in L' \sqcup\!\sqcup_{kt} A^* - L'$, donc il existe $u_0 \in L'$ tel que $u_0 \leqslant^{kt} u_1$.

- **Propriété**

  De toute chaîne L-alternée positive au seuil t, on peut obtenir pour tout $k > 0$ une chaîne de même signe et même longueur au seuil kt.

- **Propriété**

  De toute chaîne L-alternée positive au seuil t, on peut obtenir pour tout $k > 0$ une chaîne de même signe et même longueur au seuil kt.

- **Autrement dit**

  $L \notin \Sigma_1[< +MOD_t](n) \Rightarrow \forall k > 0, L \notin \Sigma_1[< +MOD_{kt}](n)$

Conclusion de la preuve

- **Théorème**
  Soit $L \in \mathcal{B}\Sigma_1[< +MOD]$, et t l'indice de stabilité du monoïde issu de L.

$$L \in \Sigma_1[< +MOD](n) \Rightarrow L \in \Sigma_1[< +MOD_t](n)$$

Conclusion de la preuve

- **Théorème**
  Soit $L \in \mathcal{B}\Sigma_1[< +MOD]$, et t l'indice de stabilité du monoïde issu de L.

$$L \in \Sigma_1[< +MOD](n) \Rightarrow L \in \Sigma_1[< +MOD_t](n)$$

  - Preuve

$$\begin{aligned} L \in \Sigma_1[< +MOD](n) &\Rightarrow \exists d. L \in \Sigma_1[< +MOD_d](n) \\ &\Rightarrow \exists d. L \in \Sigma_1[< +MOD_d + MOD_t](n) \\ &\Rightarrow \exists d. L \in \Sigma_1[< +MOD_{ppcm(d,t)}](n) \\ &\Rightarrow L \in \Sigma_1[< +MOD_t](n) \end{aligned}$$

Conclusion

- Ce que l'on a obtenu
  - La décidabilité de la hiérarchie booléenne sur $\Sigma_1[< +MOD]$
  - Une meilleure compréhension de $\Sigma_1[< +MOD]$ et de sa clôture booléenne

Conclusion

- Ce que l'on a obtenu
  - La décidabilité de la hiérarchie booléenne sur $\Sigma_1[< +MOD]$
  - Une meilleure compréhension de $\Sigma_1[< +MOD]$ et de sa clôture booléenne
- Ce que l'on veut faire ensuite
  - Étendre les résultats à d'autres signatures ($\Sigma_1[Reg]$)
  - Voire à d'autres logiques comme $\Sigma_2[<]$
  - Généraliser cette technique