Langages reconnaissables et commutations totales et partielles

Antonio Cano Gómez[1] Giovanna Guaiana[2]
Jean-Éric Pin[3]

[1]Universidad Politécnica de Valencia

[2]LITIS, Université de Rouen

[3]LIAFA, CNRS et Université Paris Diderot

11 février 2008, Paris

Sommaire

Première partie I

Commutations partielles

Commutations partielles

Soit A un alphabet fini et soit I une relation symétrique et irréflexive sur A (souvent appelée relation d'indépendance). On note $\sim_I$ la congruence sur A^* engendrée par l'ensemble

$$\{ab = ba \mid (a, b) \in I\}$$

Si L est un langage de A^*, on note $[L]_I$ la fermeture de L par $\sim_I$ (fermeture par I-commutation).

On note D la relation (dite de dépendance) $(A \times A) \setminus I$.

Commutations

Si $I = \{(a,b) \in A^2 \mid a \neq b\}$, on simplifie les notations en $\sim$ et $[L]$. Ainsi $\sim$ est la relation de commutation, $A^*/\sim\, = \mathbb{N}^A$ est le monoïde commutatif libre sur A et $[L]$ est la fermeture commutative de L.

Commutations

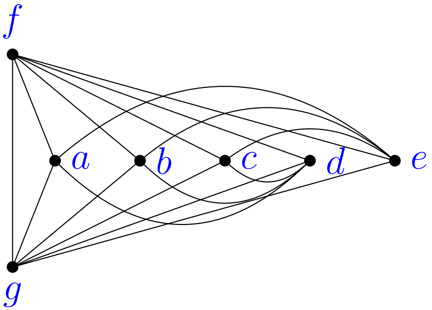
Si $I = \{(a, b) \in A^2 \mid a \neq b\}$, on simplifie les notations en $\sim$ et $[L]$. Ainsi $\sim$ est la relation de commutation, $A^*/\!\sim = \mathbb{N}^A$ est le monoïde commutatif libre sur A et $[L]$ est la fermeture commutative de L.

On peut aussi considérer des relations non symétriques, appelées semi-commutations. Ce sont des systèmes de récriture particuliers dont les règles sont de la forme $ab \to ba$, avec $a \neq b$ et $a, b \in A$.

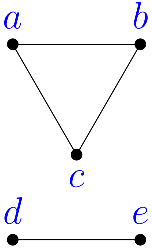
Exemple de commutation partielle

Par exemple, si $A = \{a, b, c, d, e, f, g\}$ et si I et D sont les relations ci-dessous, on a
$A^*/{\sim_I} = \{a, b, c\}^* \times \{d, e\}^* \times \{f\}^* \times \{g\}^*$.

I



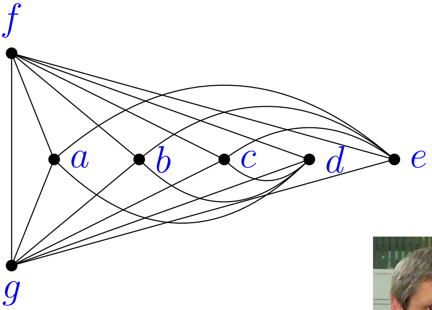
D



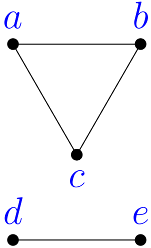
Exemple de commutation partielle

Par exemple, si $A = \{a, b, c, d, e, f, g\}$ et si I et D sont les relations ci-dessous, on a $A^*/{\sim_I} = \{a, b, c\}^* \times \{d, e\}^* \times \{f\}^* \times \{g\}^*$.

I



D



Utilisez GasTEX !

Deux types de questions

- Étant donné une commutation partielle I, quels sont les langages reconnaissables L tels que $[L]_I$ soit reconnaissable ?

Deux types de questions

- Étant donné une commutation partielle I, quels sont les langages reconnaissables L tels que $[L]_I$ soit reconnaissable ?

- Y-a-t-il des classes robustes de langages reconnaissables fermées par l'opération $L \to [L]_I$?

Robuste $=$ fermée par pour des opérations usuelles telles qu'union, intersection, complément, produit, étoile, mélange, inverse de morphismes, morphismes, quotients, etc.

Deuxième partie II

État de l'art

Fermeture commutative

Théorème (Ginsburg-Spanier 66, Gohon 85)

On peut décider si la fermeture commutative $[L]$ *d'un langage reconnaissable* L *est reconnaissable.*

Le langage $[(ab)^*] = \{u \in \{a,b\}^* \mid |u|_a = |u|_b\}$ n'est pas reconnaissable. Cependant $[(ab)^* + A^*aaA^* + A^*bbA^*]$ est reconnaissable.

Fermeture par commutation partielle

Théorème (Sakarovitch 1992)

Le problème de savoir si $[L]_I$ est reconnaissable est décidable ssi I est une relation transitive.

Polynômes de langages

Soit $\mathcal{L}$ une classe de langages reconnaissables. On note $\mathrm{Pol}(\mathcal{L})$ fermeture polynomiale de $\mathcal{L}$, formée des unions finies de produits de la forme

$$L_0 a_1 L_1 \cdots a_k L_k$$

avec $a_1, \ldots, a_k \in A$ et $L_0, \ldots, L_k \in \mathcal{L}$.

Exemples

Si $\mathcal{I}(A^*) = \{\emptyset, A^*\}$, Pol$(\mathcal{I})$ est la classe des unions finies de produits de la forme $A^*a_1A^* \cdots a_kA^*$ avec $a_1, \ldots, a_k \in A$. On note $\mathcal{J}$ la fermeture booléenne de Pol$(\mathcal{I})$ (langages testables par morceaux).

Les unions finies de langages de la forme

$$A_0^*a_1A_1^* \cdots a_kA_k^*$$

avec $A_i \subseteq A$, ont été baptisés APC (Alphabetic Pattern Constraints) par Bouajjani, Muscholl et Touili. Autres notations : Pol$(\mathcal{J})$, $\mathcal{V}_{3/2}$.

Deuxième question

Théorème (Guaiana, Restivo, Salemi 00, 04)

Les classes $\mathit{Pol}(\mathcal{I})$ *et* $\mathcal{J}$ *sont fermées par commutation.*

Théorème (Guaiana, Restivo, Salemi 00, 04, Bouajjani, Muscholl, Touili 01, 07)

La classe $\mathit{Pol}(\mathcal{I})$ *(APC) est fermée par semi-commutation partielle.*

Deuxième question (suite)

Un langage L est commutatif si $[L] = L$. On note $\mathcal{C}om$ la classe des langages commutatifs.

Théorème (Cécé, Héam, Mainier 2006,2007)

La classe Pol$(\mathcal{C}om)$ est fermée par semi-commutation partielle.

Langages sans étoile

Théorème (Muscholl, Petersen 1996)

Soit L un langage sans-étoile de A^*. Si D est transitive, alors $[L]_I$ est soit sans-étoile, soit non reconnaissable. C'est faux si D est non transitive.

Théorème (Muscholl, Petersen 1996)

On peut décider si la fermeture $[L]_I$ d'un langage sans-étoile L est sans-étoile ssi I est transitive.

Troisième partie III

Fermeture commutative

Langages à groupe

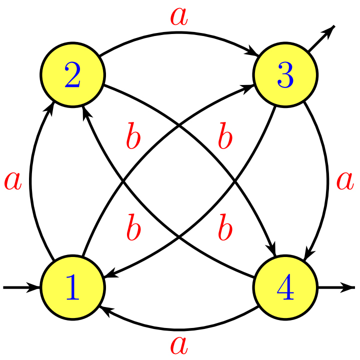
Un automate de permutations est un automate déterministe dans lequel chaque lettre définit une permutation sur l'ensemble des états.

Definition

Un langage à groupe est un langage reconnu par un groupe fini ou, de façon équivalente, par un automate de permutations.

On note $\mathcal{G}$ la classe des langages à groupe.

Un automate de permutations



Fermeture commutative des langages à groupe

Théorème

La fermeture commutative d'un langage à groupe est un langage reconnaissable.

Preuve. Soit L un langage à groupe, soit $\pi : A^* \to G$ son morphisme syntactique et soit $n = |G|$. Il suffit de prouver qu'il existe un entier N tel que, pour chaque lettre a, que $a^N \sim_{[L]} a^{N+n}$.

Comme G est un groupe, $\pi(a^n) = \pi(a)^n = 1$. Donc si $xa^N y \in [L]$, on a $xa^N y \sim u$ avec $u \in L$, d'où $xa^{N+n} y \sim ua^n$ avec $ua^n \in L$. Donc $xa^{N+n} y \in [L]$.

Fin de la preuve

Supposons $xa^{N+n}y \in [L]$. Alors $xa^{N+n}y \sim u \in L$. Donc $u = u_0 a u_1 a \cdots u_{N-1} a u_N a u_{N+1}$. Si on choisit N assez grand (Ramsey), on peut trouver une suite $0 \leqslant i_0 < i_1 < \ldots < i_n \leqslant N$ telle que

$$\pi(u_{i_0} a \cdots u_{i_1 - 1} \underline{a}) = \cdots = \pi(u_{i_n - 1} a \cdots u_{i_n} \underline{a}) = 1$$
$$\pi(u_{i_0} a \cdots a u_{i_1 - 1}) = \cdots = \pi(u_{i_n - 1} a \cdots a u_{i_n}) = \pi(a)^{-1}$$

Soit u' le mot déduit de u en supprimant les n lettres a. On a $u' \sim xa^N y$ et $u' \in L$ car

$$\pi\big((u_{i_0} a \cdots a u_{i_1 - 1}) \cdots (u_{i_n - 1} \cdots a u_{i_n})\big) = \pi(a)^{-n} = 1$$

Polynômes de langages à groupe ($\mathsf{Pol}(\mathcal{G})$)

Théorème

La fermeture commutative d'un langage de $\mathsf{Pol}(\mathcal{G})$ *est aussi dans* $\mathsf{Pol}(\mathcal{G})$.

On utilise le résultat suivant :

Théorème (JEP 96)

Si L *est dans* $\mathsf{Pol}(\mathcal{G})$, *il existe un morphisme surjectif* $\pi : A^* \to G$, *où* G *est un groupe fini, tel que* L *soit union finie de langages de la forme* $Ra_1R \cdots Ra_nR$, *avec* $R = \pi^{-1}(1)$.

Fermeture commutative des langages de Pol($\mathcal{G}$)

On considère le système d'insertion dont les règles sont $1 \to r$, avec $r \in R$. On écrit $u \to_R v$ si $u = u_0u_1$ and $v = u_0ru_1$ avec $r \in R$. Enfin, on note $\xrightarrow{*}_R$ la fermeture réflexive et transitive de $\to_R$.

Théorème (Buchler-Ehrenfeucht-Haussler 85)

La relation $\xrightarrow{*}_R$ est un beau préordre (wqo).

Corollaire

Pour tout langage L, $[L]_{\xrightarrow{*}_R}$ est dans Pol($\mathcal{G}$).

Polynômes de langages à groupe

Théorème

La fermeture commutative d'un langage de Pol($\mathcal{G}$) est aussi dans Pol($\mathcal{G}$).

Preuve. Il suffit de prouver le résultat pour $L = Ra_1R \cdots Ra_nR$. On vérifie pour cela que $[L]$ est fermé pour $\xrightarrow{*}_R$, ce qui est assez facile.

Période d'un langage reconnaissable

La période d'un langage reconnaissable est le plus petit entier p tel qu'il existe un entier n tel que, pour tout $u \in A^*$, $u^n \sim_L u^{n+p}$.

Par exemple un langage de période 1 est un langage sans étoile (Schützenberger 65).

La variété $\mathcal{W}$

Une variété positive de langages est une classe de langages reconnaissables fermée par union finie, intersection finie, quotients ($L \to u^{-1}L, Lu^{-1}$) et inverse de morphismes.

Théorème (Cano Gómez-JEP, 2002)

Il existe une variété positive maximale, notée $\mathcal{W}$, qui ne contient pas le langage $(ab)^*$. Cette variété est décidable.

La variété $\mathcal{W}$ est très robuste. . .

- $\mathcal{W}$ est l'unique variété positive maximale propre fermée par mélange (= shuffle),
- $\mathcal{W}$ est fermée par produit,
- $\mathcal{W}$ est fermée par morphisme alphabétique,
- $\mathcal{W}$ contient les langages à groupe, les langages commutatifs et leur fermeture polynomiale.
- $\mathcal{W}$ contient les langages dont le monoïde syntactique est dans la variété **DS**.

Exemple. Par définition, $(ab)^* \notin \mathcal{W}$. Cependant $(ab)^* + A^*aaA^* + A^*bbA^* \in \mathcal{W}$.

La variété de monoïdes ordonnés **W**

Théorème (Cano Gómez-JEP, 2002)

Un monoïde ordonné M appartient à **W** *ssi, pour tout couple (a, b) d'éléments de M tels que $aba = a$ et $bab = b$, pour tout élément z de l'idéal minimal du sous-monoïde de M engendré par a et b, on a $(abzab)^\omega \leqslant ab$.*

Fermeture commutative des langages de $\mathcal{W}$

Théorème

Si L est un langage de $\mathcal{W}(A^*)$, $[L]$ est reconnaissable et sa période divise celle de L.

Preuve. Algébrique et assez délicate.

Corollaire

La classe $\mathcal{W}$ est fermée par commutation. La classe des langages sans étoile de $\mathcal{W}$ est fermée par commutation.

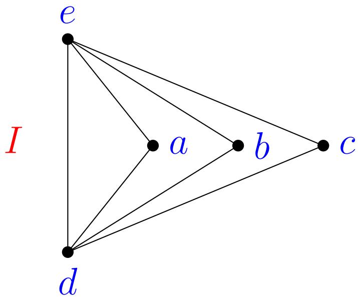
Quatrième partie IV

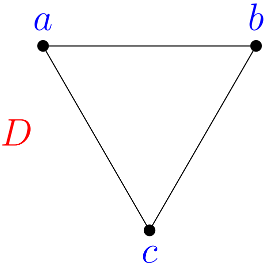
Commutations partielles

Le cas où D est une clique

Théorème

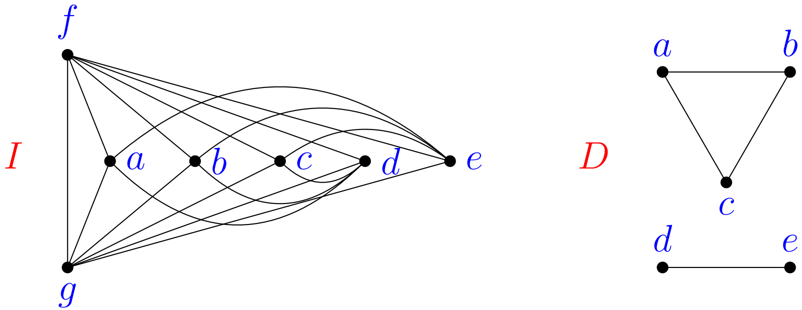
Si L est un polynôme de langages à groupes, $[L]_I$ est un polynôme de langages à groupes.





Preuve. Raffinement de la deuxième preuve.

Le cas où D est transitive



Supposons que $A^*/{\sim_I} = A_1^* \times \cdots \times A_k^*$. On note $\pi_j : A^* \to A_j^*$ la projection naturelle et $\pi_I : A^* \to A_1^* \times \cdots \times A_k^*$ le morphisme défini par

$$\pi_I(u) = \{\pi_1(u), \ldots, \pi_k(u)\}$$

Une formule explicite (D transitive)

Théorème

Si $\pi_I(L)$ est une partie reconnaissable de $A_1^* \times \cdots \times A_k^*$, alors $[L]_I$ est reconnaissable.

Si

$$\pi_I(L) = \bigcup_{1 \leqslant i \leqslant n} L_{i,1} \times \cdots \times L_{i,k}$$

Alors

$$[L]_I = \bigcup_{1 \leqslant i \leqslant n} L_{i,1} \sqcup\!\sqcup \cdots \sqcup\!\sqcup L_{i,k}$$

Polynômes de langages à groupe (D transitive)

Théorème

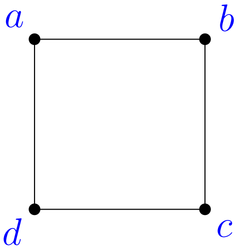
Si L est un langage à groupe, $\pi_I(L)$ est une partie reconnaissable de $A_1^* \times \cdots \times A_k^*$.

Théorème

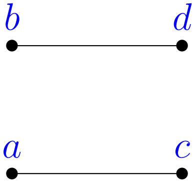
Si L est un polynôme de langages à groupes, $[L]_I$ est un polynôme de langages à groupes.

Le résultat ne s'étend pas à $\mathcal{W}$ (D transitive)

I



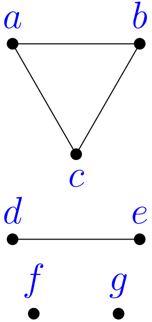
D



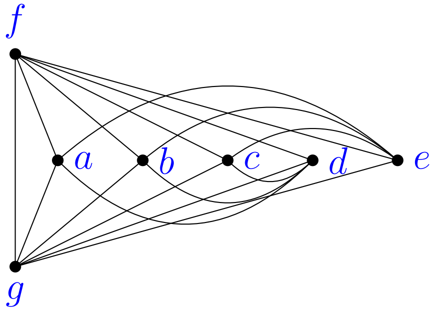
Le langage $L = (abcd)^* + A^*aaA^* + A^*bbA^* + A^*ccA^* + A^*ddA^* + A^*ababA^* + A^*bcbcA^* + A^*cdcdA^* + A^*dadaA^*$ est dans $\mathcal{W}$, mais $[L]_I$ n'est pas reconnaissable.

Le cas où I est transitive

I



D



Dans ce cas, $A^*/{\sim_I}$ est égal à un produit libre $\mathbb{N}^{A_1} * \cdots * \mathbb{N}^{A_k}$. Ici, $\{\{a, b, c\}, \{d, e\}, \{f\}, \{g\}\}$.

Le cas I transitive

Théorème

Si L est un langage de $\mathcal{W}$ et si I est transitive, alors $[L]_I$ est un langage reconnaissable.

Preuve. Soit $\mathcal{A} = (Q, A, \cdot, q_0, F)$ l'automate minimal de L. On ordonne les états par $p \leqslant q$ ssi,

$$\text{pour tout } u \in A^*,\ q{\cdot}u \in F \Rightarrow p{\cdot}u \in F$$

Le cas I transitive (suite)

On construit un automate $\mathcal{B}$ avec le même ensemble d'états Q et des transitions

$$p \xrightarrow{R_{p,q}} q$$

avec $R_{p,q} = \bigcup_{1 \leqslant i \leqslant k} \big[\{u \in A_i^* \mid p{\cdot}u \leqslant q\}\big]$.

On démontre ensuite que $\mathcal{B}$ reconnaît $[L]_I$. Par ailleurs, chaque langage $\{u \in A_i^* \mid p{\cdot}u \leqslant q\}$ est dans $\mathcal{W}$ et donc chaque $R_{p,q}$ est reconnaissable.

$\mathcal{B}$ reconnaît $[L]_I$

Soit $u \in [L]_I$. On factorise u en $u_1 \cdots u_n$ sur le modèle $(acba)(de)(bcba)(g)(de)(f)(g)$. Il existe alors des mots $v_1, \ldots, v_n$ tels que $u_1 \sim v_1, \ldots,$ $u_n \sim v_n$ et $v_1 \cdots v_n \in L$.

Soient $q_1 = q_0 \cdot v_1$, $q_2 = q_1 \cdot v_2$, $\ldots$, $q_n = q_{n-1} \cdot v_n$.

$$\to q_0 \xrightarrow{v_1} q_1 \xrightarrow{v_2} q_2 \quad \cdots \quad q_{n-1} \xrightarrow{v_n} q_n \to$$

Comme $v_1 \cdots v_n \in L$, q_n est un état final. De plus, $u_1 \in R_{q_0,q_1}$, $\ldots$, $u_n \in R_{q_{n-1},q_n}$ et donc u est accepté par $\mathcal{B}$.

$\mathcal{B}$ reconnaît $[L]_I$ (Fin)

Soit u accepté par $\mathcal{B}$ et soit

$$\to q_0 \xrightarrow{u_1} q_1 \xrightarrow{u_2} q_2 \quad \cdots \quad q_{n-1} \xrightarrow{u_n} q_n \to$$

un chemin réussi de $\mathcal{B}$ d'étiquette u. Donc q_n est final et les lettres de chaque u_i commutent. De plus, il existe des mots $v_1, \ldots, v_n$ tels que $u_1 \sim v_1$, $\ldots$, $u_n \sim v_n$ et $q_0 \cdot v_1 \leqslant q_1$, $q_1 \cdot v_2 \leqslant q_2$, $\ldots$, $q_{n-1} \cdot v_n \leqslant q_n$. Il en résulte que $q_0 \cdot v_1 \cdots v_n \leqslant q_n$.

Mais d'après la définition de $\leqslant$, la condition $q_n \in F$ entraîne $q_0 \cdot v_1 \cdots v_n \in F$ et donc $v_1 \cdots v_n \in L$. Par conséquent $u \sim_I v$ et $u \in [L]_I$.

Une extension

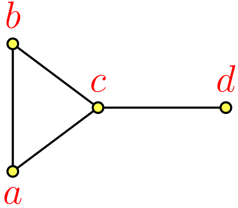
Soient $I_1, \ldots, I_k$ les composantes connexes du graphe (A, I). On note D_j le complément de I_j. Dans le cas de $\mathrm{Pol}(\mathcal{G})$, la preuve s'étend au cas où chaque D_j est transitif.

Théorème

On suppose chaque D_j est transitif. Alors si L est un polynôme de langages à groupe, $[L]_I$ est reconnaissable.

Un peu de théorie des graphes

Soient P_4 et paw les graphes ci-dessous :



Théorème

Les composantes D_j sont transitives ssi (A, I) n'a aucun sous-graphe égal à P_4 ou à paw.

Reformulation

Théorème

On suppose que le graphe (A, I) ne possède aucun sous-graphe du graphe (A, I) égal à P_4 ou à paw. Alors si L est un polynôme de langages à groupe, $[L]_I$ est reconnaissable.

Problèmes ouverts

- Si I est transitive et si L est dans $\mathcal{W}$, est-ce que $[L]_I$ est dans $\mathcal{W}$?
- Si L est un langage à groupe, est-ce que $[L]_I$ est toujours reconnaissable ?
- Est-ce que $\mathrm{Pol}(\mathcal{G})$ est fermé par commutations partielles ?
- Cas de la fermeture polynomiale des langages commutatifs ou à groupe ?
- Extension à d'autres systèmes de réécriture (par exemple $abb = bba$) ?