ENSEIRB-MATMECA- Section Informatique, 2ième année
Calculabilité, 2020/2021

CALCULABILITÉ

DM- à rendre avant le 19/03/2021, à midi [1]

Indications : Chaque partie dépend des parties précédentes.
On peut *admettre* une question (voire même toute une partie) et utiliser les résultats de cette question (ou partie) dans la suite du problème.
Tous moyens d'investigation autorisés. Rédaction finale par groupes de quatre.
Sujet : Ce problème est centré sur le problème de correspondance de Post [2].
Nous décrirons le problème (partie I) et montrerons qu'il est indécidable (Partie IV), en passant par un problème sur la réécriture de mots (partie II) et par le problème de l'acceptation pour les machines de Turing (partie III).
Deux extensions (facultatives) : Nous exhiberons une instance du (PCP) qui a une solution extrêmement *longue* (partie V), puis nous étudierons des variantes *infinies* du problème de réécriture et du (PCP) (partie VI).

Partie I

Problème de Post.

Le problème de correspondance de Post (PCP) est le suivant :

- Instance : une suite $I = ((u_1, v_1), \dots (u_n, v_n))$ de couples de mots finis sur un alphabet Σ

- Question : existe-t-il un entier k et une suite $i_1, \dots, i_k$ à valeurs dans $[1, n]$ tels que $u_{i_1} u_{i_2} \cdots u_{i_k} = v_{i_1} v_{i_2} \cdots v_{i_k}$?

On dénote par $X = \{x_1, x_2, \dots, x_n\}$ un alphabet de cardinal n et on note φ, ψ les homomorphismes de X^* dans Σ^* tels que

$$\forall i \in [1, n],\ \varphi(x_i) = u_i,\ \ \psi(x_i) = v_i. \tag{1}$$

1- Vérifier que l'instance I de (PCP) a la réponse "oui" ssi

$$\exists w \in X^+, \varphi(w) = \psi(w).$$

2- Soit $S(I) = \{w \in X^* \mid \varphi(w) = \psi(w)\}$ (c'est l'ensemble des solutions au PCP, augmenté du mot vide).

1. v3 du 11/02/22 issue des versions v2 du 15/03/21, v1 du 04/03/21 et des critiques pertinentes des étudiants

2. Du nom de son auteur, Emil Post : *A variant of a recursively unsolvable problem.* Bulletin of the American Mathematical Society 52, 1946.

$S(I)$ est-il clos par produit ? par facteur ? S'agit-il d'un sous-monoide de X^* ?

On appelle solution *atomique* de I une solution $w \in X^+$ telle que, pour toute décomposition $w = w_1 \cdot w_2$, avec $w_1, w_2 \in \mathrm{S}(I)$, $w_1 = \varepsilon$ ou $w_2 = \varepsilon$. On note $\mathrm{A}(I)$ l'ensemble des solutions atomiques de I.
3- Montrer que $\mathrm{S}(I) = \mathrm{A}(I)^*$.
4- Résoudre le problème PCP sur les instances suivantes :
$I_1 = ((bbb, bb), (abb, babbb))$
$I_2 = ((ba, bab, (abb, bb), (bab, abb))$
$I_3 = ((\#\#x, \#\#xaxaxbxbxbx\#), (\#xbxbxbxaxax\#\#, x\#\#), (ax, xa), (bx, xb), (\#x, x\#), (axbx, xbxa))$
Aide : on pourra revenir sur I_3 après avoir traité les parties II et IV.

Partie II

Systèmes semi-Thuéiens.

Un système de réécriture sur un alphabet A est une partie $S \subseteq A^* \times A^*$. La relation binaire de *dérivation* immédiate $\to_S$ est définie par :
$\forall u, v \in A^*, u \to_S v$ si et seulement si

$$\exists \alpha \in A^*, \exists \beta \in A^*, \exists (\ell, r) \in S, \;\; u = \alpha \ell \beta, v = \alpha r \beta.$$

Pour tout entier $n \geq 0$, une dérivation *d'ordre* n est une suite de mots

$$D = (u_0, u_1, \ldots, u_i, u_{i+1}, \ldots u_n)$$

telle que $\forall i \in [0, n-1], u_i \to_S u_{i+1}$.
NB : Une dérivation d'ordre 0 est donc une suite réduite à un seul mot (u_0).
On dit que D va de u_0 à u_n. On note $u \to_S^n v$ ssi il existe une dérivation D d'ordre n qui va de u en v. On note

$$\to_S^* \; := \bigcup_{n \geq 0} \to_S^n .$$

Le problème de l'accessibilité pour les systèmes semi-Thuéiens [3] (ACC-ST) est le suivant :

— Instance : un triplet (S, u, v) formé d'un système de réécriture fini S sur un alphabet A et deux mots $u, v \in A^*$.

3. les systèmes de réécriture de mots sont aussi appelés systèmes semi-Thuéiens, du nom du mathématicien Norvégien, Axel Thue, qui les a étudiés dès 1914, avant que la notion d'algorithme ne soit définie formellement

— Question : $u \to_S^* v$?
c'est à dire est-ce-qu'il existe une dérivation allant de u à v dans le système semi-Thuéien S.

1- Montrer que pour les systèmes S qui conservent la longueur (i.e. $\forall(\ell, r) \in S,\ \ |\ell| = |r|$), le problème (ACC-ST) est *décidable*.
2- Supposons qu'il existe un ordre total $\sqsubseteq$ sur les mots de A^*, qui est compatible avec le produit i.e. tel que

$$\forall \alpha, u, v, \beta \in A^*, u \sqsubseteq v \Rightarrow \alpha \cdot u \cdot \beta \sqsubseteq \alpha \cdot v \cdot \beta$$

et tel que les ensembles ordonnés $(\mathbb{N}, \leq)$ et $(A^*, \sqsubseteq)$ soient isomorphes. Un système S est dit *croissant* pour l'ordre $\sqsubseteq$ si, $\forall(\ell, r) \in S,\ \ \ell \sqsubseteq r$. Montrer que pour les systèmes S croissants pour un ordre vérifiant les hypothèses ci-dessus, le problème (ACC-ST) est décidable.
3- En déduire que le problème (ACC-ST) est décidable pour les systèmes S qui n'ont qu'une règle.
4- Résoudre le problème (ACC-ST) sur les instances suivantes :

$S_1 = \quad \{(ab, ba)\}, u_1 = aabbb,\ \ v_1 = bbbaa$

$S_2 = \quad \{(ab, ba), (ab, aab), (aa, ba)\}, u_2 = aab,\ \ v_2 = ababa$

$S_3 = \quad \{(x0, x1),\ \ (x1, x0)\} \cup \{(x0^n10, x0^n11) \mid 0 \leq n \leq 5\} \cup \{(x0^n11, x0^n10) \mid 0 \leq n \leq 5\},$
$u_3 = x1111111,\ \ v_3 = x0000000$

$S_4 = \quad \{(aba, \varepsilon),\ \ (\varepsilon, aba)\},\ \ u_4 = a^{4042}b^{2021},\ \ v_4 = (ab)^{2021}a^{2021}.$

Aide : pour S_3 penser au casse-tête nommé "baguenaudier".

Partie III

Machines de Turing versus Systèmes semi-Thuéiens.

Soit $\mathcal{M} = (\Sigma, Q, q_-, Q_+, \delta)$ une machine de Turing à une bande (avec $\Sigma = \{\triangleright, \square\} \cup \Gamma$ où $\triangleright \neq \square$ et $\{\triangleright, \square\} \cap \Gamma = \emptyset$). L'état q_- est l'état de départ et Q_+ est l'ensemble des états d'acceptation de la machine. Une *configuration* de $\mathcal{M}$ est un mot infini vers la droite $c \in (\Sigma \cup Q)^\omega$ de la forme :

$$c = uqv\square^\omega$$

où $u, v \in \Sigma^*, q \in Q$ et $u \cdot v \in \triangleright(\Sigma \setminus \{\triangleright\})^*$.

L'idée intuitive est que la machine a un contenu de ruban qui est $u \cdot v \cdot \Box^\omega$, est dans l'état q et sa tête de lecture pointe sur la première position de $v \cdot \Box^\omega$.

On suppose que $\mathcal{M}$ n'a que des directions L, R dans ses instructions. La relation de mouvement $\vdash_{\mathcal{M}}$ sur les configurations de $\mathcal{M}$ est définie par :

$c \vdash_{\mathcal{M}} c'$ ssi l'un des deux cas suivants se produit :

$$c = \alpha q x \beta, c' = \alpha y q' \beta \text{ où } \quad \alpha \in \Sigma^*, \beta \in \Sigma^\omega, \quad \text{et } (q, x, q', y, R) \in \delta$$

$$c = \alpha z q x \beta, c' = \alpha q' z y \beta \text{ où } \quad \alpha \in \Sigma^*, z \in \Sigma, \beta \in \Sigma^\omega, \quad \text{et } (q, x, q', y, L) \in \delta.$$

Introduisons un nouveau symbole $\# \notin \Sigma \cup Q$ et posons $\hat{\Sigma} = \Sigma \cup Q \cup \{\#\}$. Une *description instantanée* de la machine $\mathcal{M}$ est un mot fini $d \in \hat{\Sigma}^*$ de la forme :

$$d = uqv\#$$

où $u, v \in \Sigma^*, q \in Q$ et $u \cdot v \in \triangleright(\Sigma \setminus \{\triangleright\})^*$.

On note $\mathsf{CONF}(\mathcal{M})$ (resp. $\mathsf{DI}(\mathcal{M})$)) l'ensemble des *configurations* (resp. *Descriptions Instantannées*) de la machine $\mathcal{M}$. Soit $\mathsf{C} : \mathsf{DI}(\mathcal{M})) \to \mathsf{CONF}(\mathcal{M})$ l'application qui remplace le symbole $\#$ par une infinité de blancs sur la droite :

$$\mathsf{C}(uqv\#) = uqv \cdot \Box^\omega.$$

1- L'application C est-elle injective ? surjective ?

2- Donner un système semi-Thuéien $S_{\mathcal{M}}$ sur Σ^* tel que, pour tous $d_1, d_2 \in \mathsf{DI}(\mathcal{M})$ on ait :

$$d_1 \to_{S_{\mathcal{M}}}^* d_2 \Rightarrow \mathsf{C}(d_1) \vdash_{\mathcal{M}}^* \mathsf{C}(d_2)$$

et pour tous $c_1, c_2 \in \mathsf{CONF}(\mathcal{M})$ et $d_1 \in \mathsf{DI}(\mathcal{M})$ on ait :

$$(\mathsf{C}(d_1) = c_1 \text{ et } c_1 \vdash_{\mathcal{M}}^* c_2) \Rightarrow (\exists d_2 \in \mathsf{DI}(\mathcal{M}), d_1 \to_{S_{\mathcal{M}}}^* d_2 \text{ et } \mathsf{C}(d_2) = c_2).$$

En mots : le système semi-Thuéien $S_{\mathcal{M}}$ *simule*, sur les descriptions instantanées, les mouvements de $\mathcal{M}$ sur les configurations(voir la figure 1). [4]

3- Montrer que la machine $\mathcal{M}$ accepte un mot $u \in \Gamma^*$ ssi,

$$\exists v \in \Sigma^* Q_+ \Sigma^*, \quad q_- \triangleright u\# \to_{S_{\mathcal{M}}}^* v\#. \tag{2}$$

4- On considère l'alphabet $\Sigma' := \Sigma \cup Q \cup \{\#, \bar{\#}\}$ obtenu en ajoutant un nouveau symbole $\bar{\#}$ à $\hat{\Sigma}$. Construire un système semi-Thuéien $T_{\mathcal{M}}$ sur l'alphabet Σ' tel que, pour tout mot $u \in \Gamma^*$, (2) est vrai ssi

$$q_- \triangleright u\# \to_{T_{\mathcal{M}}}^* \bar{\#}.$$

4. techniquement : C est une *bisimulation* fonctionnelle

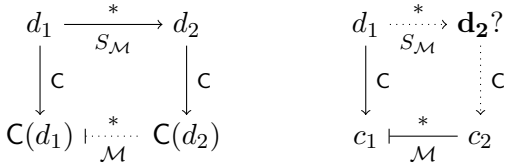

FIGURE 1 – le système semi-Thuéien (bi)-simule la machine

5- Donner une réduction du problème (ACC-MT) (le problème de l'acceptation pour les machines de Turing) au problème (ACC-ST).
6- Montrer que, même en se restreignant aux systèmes semi-Thuéiens sur l'alphabet à 2 lettres $\{a, b\}$, le problème (ACC-ST) reste indécidable.

Partie IV

Systèmes semi-Thuéiens versus Problème de Post.

Soit $S = \{(\ell_i, r_i) \mid 1 \leq i \leq n\}$ un système de réécriture sur $\{a, b\}^*$ et soient $u, v \in \{a, b\}^*$. On définit $\Phi(S, u, v)$ comme l'instance suivante du (PCP) :

$$\begin{array}{c}((\#\#x, \#\#xu^x\#), (\#xv^x\#\#, x\#\#), \quad (ax, xa), (bx, xb), (\#x, x\#), \\ (\ell_1^x,\ {}^x r_1), \ldots, (\ell_n^x,\ {}^x r_n))\end{array} \tag{3}$$

où $x, \#$ sont des nouvelles lettres ($x \notin \{a, b\}, \# \notin \{a, b\}, x \neq \#$) et les exposants ont la signification suivante : pour tout mot $w = x_1 x_2 \ldots x_m$ (où $m \geq 1$)

$$w^x = x_1 x x_2 x \ldots x_m x, \quad {}^x w = x x_1 x x_2 \ldots x x_m, \ \ \varepsilon^x = {}^x \varepsilon = \varepsilon.$$

1- Montrer que, si $u = u_0 \to_S u_1 \ldots \to_S u_p = v$, alors $\Phi(S, u, v)$ a une solution utilisant p occurences de couples de l'ensemble $\{(\ell_i^x,\ {}^x r_i) \mid 1 \leq i \leq n\}$.

On renomme $D, F, L_a, L_b, L_\#, R_1, \ldots, R_n$ les lettres de X (prises dans l'ordre de la définition (3)), et on continue de noter φ, ψ les homomorphismes définis en (1). Ainsi :

$$\begin{array}{lll} \varphi(D) = \#\#x, & \psi(D) = \#\#xu^x\#, & \varphi(F) = \#xv^x\#\#,\ \psi(F) = x\#\#, \\ \varphi(L_a) = ax, & \psi(L_a) = xa,\ \ \varphi(L_b) = bx, & \psi(L_b) = xb,\ \ \varphi(L_\#) = \#x,\ \psi(L_\#) = x\# \\ \varphi(R_1) = \ell_1^x, & \psi(R_1) = {}^x r_1, \ldots, \varphi(R_i) = \ell_i^x, & \psi(R_i) = {}^x r_i, \ldots, \varphi(R_n) = \ell_n^x,\ \psi(R_n) = {}^x r_n. \end{array}$$

2- Montrer que si $H \in \{D, F, L_a, L_b, L_\#, R_1, \ldots, R_n\}^*$ est une solution atomique, alors H est de la forme

$$H = D \cdot H_0 \cdot F$$

avec $H_0 \in \{L_a, L_b, L_\#, R_1, \ldots, R_n\}^*$ et

$$\varphi(H_0)\#\ {}^x v = u^x \# \psi(H_0)$$

3- Montrer que, si $H \in \{L_a, L_b, R_1, \ldots, R_n\}^*$ alors $\exists u_0, v_0 \in \{a, b\}^*$ tels que :

$$\varphi(H) = u_0^x, \quad \psi(H) = {}^x v_0, \quad u_0 \to_S^p v_0 \text{ où } p = |H|_{R_1,\ldots,R_n}.$$

4- Montrer que, pour tous mots $H \in \{L_a, L_b, L_\#, R_1, \ldots, R_n\}^*, u_1 \in \{a, b\}^*, v_1 \in \{a, b\}^*$,

$$\varphi(H)\#\ {}^x v_1 = u_1^x \# \psi(H) \Rightarrow [u_1 \to_S^p v_1 \text{ où } p = |H|_{R_1,\ldots,R_n}].$$

Aide : raisonner par récurrence sur $|H|_{L_\#}$.
5- Montrer que $u \to_S^* v$ si et seulement si $\Phi(S, u, v)$ a une solution.
6- Conclure que Φ est une réduction du problème (ACC-ST) au problème (PCP) et que (PCP) est indécidable.
7- Pouvez-vous maintenant résoudre (PCP) sur l'instance I_3 de la partie I, question 4 ?

Partie V

Problème de Post à longue solution.

Un système semi-Thuéien est dit *orthogonal* si ses membres gauches n'ont pas de chevauchement et chaque membre gauche n'a qu'un membre droit associé : pour toutes règles $(\ell, r), (\ell', r')$ et tous mots α, m, β

$$(\ell = \alpha \cdot m \text{ et } \ell' = m \cdot \beta) \quad \Rightarrow \quad (m = \varepsilon \text{ ou } (\alpha = \beta = \varepsilon \text{ et } r = r')).$$

(voir la figure 2).
1- Montrer que, si S est orthogonal et $u \to_S v, u \to_S w$ alors il existe $w' \in A^*$ et $p', q' \in \mathbb{N}$ tels que

$$0 \le p' \le 1, \quad 0 \le q' \le 1, \quad 1 + q' = 1 + p', \quad v \to_S^{q'} w', w \to_S^{p'} w'$$

2- Montrer que si S est orthogonal et $u \to_S^p v, u \to_S^q w$ alors il existe w' et p', q' tels que

$$0 \le p' \le p, \quad 0 \le q' \le q, \quad p + q' = q + p', \quad v \to_S^{q'} w', w \to_S^{p'} w'$$

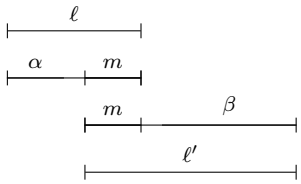

FIGURE 2 – Orthogonalité de S : configuration interdite, sauf si $\ell = \ell'$ et $r = r'$

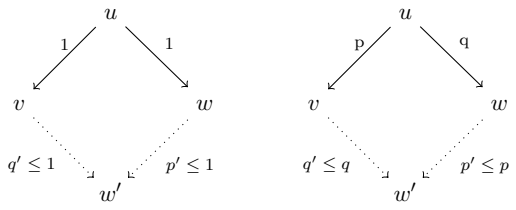

FIGURE 3 – Confluence des systèmes orthogonaux.

(voir la figure 3).
Aide : on pourra raisonner par récurrence sur $p + q$.

On appelle mot *irréductible* pour S, tout mot u tel que, il n'existe aucun mot v tel que $u \to_S v$.
3- Montrer que si S est orthogonal et $u \to_S^p v$, $u \to_S^q w$ et v, w sont irréductibles, alors

$$v = w \text{ et } p = q.$$

On considère le système de réécriture semi-Thuéien S sur l'alphabet $A := \{a, b, c, D, F\}$ défini par l'ensemble des règles :

$$ab \to bba, \quad cb \to bcc, \quad Db \to D \quad aF \to F, \quad cF \to F.$$

4-Vérifier que S est orthogonal.
5- Montrer que, pour tout entier $n \geq 0$:

$$a^n b \to_S^* b^{2^n} a^n \qquad cb^n \to_S^* b^n c^{2^n}$$

6- En déduire que

$$Dca^n bF \to_S^* Db^{2^n} c^{2^{2^n}} a^n F \to_S^* DF$$

7- Décrire une dérivation $Dca^nbF \to_S^* DF$ et donner son nombre d'applications de la règle $cF \to F$.
8- Montrer que *toute* dérivation de Dca^nbF en DF a une longueur $\geq 2^{2^n}$.
9- Donner explicitement un (PCP) dont la plus petite solution soit de longueur $\geq 2^{2^8}$.
10- Le grand professeur Cosinus (qui est parfois distrait) a construit un ordinateur égalant le super-ordinateur japonais Fujitsu/ARM (puissance de 418 Pflops, 7300 000 processeurs) afin d'énumérer la suite des indices d'une solution de l'instance du (PCP) de la question 9. Combien de temps (exprimé en milliards d'années) prendra ce calcul ? sera-t-il achevé lorsque le soleil explosera et se transformera en géante rouge ?

Partie VI

Problème de Post infinitaire.

Revenons vers les systèmes de réécriture.
Soit S un sytème de réécriture de mots. On appelle dérivation infinie (pour S) toute suite de mots

$$D = (u_0, u_1, \ldots, u_i, u_{i+1}, \ldots u_n, \ldots)$$

telle que $\forall i \in \mathbb{N}, u_i \to_S u_{i+1}$. On note $u \to_S^\infty$ le fait qu'il existe une dérivation infinie partant de u.

Le problème de la *terminaison* pour les systèmes semi-Thuéiens (TERM-ST) est le suivant :

- Instance : un couple (S, u) formé d'un système de réécriture fini S sur un alphabet A et un mot $u \in A^*$.
- Question : $\neg(u \to_S^\infty)$?
  c'est à dire, toute dérivation partant de u est finie.

1-Résoudre le problème (TERM-ST) sur les instances suivantes :

$$\begin{aligned}
S_1 = &\quad \{(ab, bba)\}, u_1 = abb.\\
S_2 = &\quad \{(ab, bba)\}, u_2 = aba.\\
S_3 = &\quad \{(bc, dc), (bd, db), (ad, abb)\}, \quad u_3 = abc.\\
S_4 = &\quad \{(bad, dadcbabb), \quad (bd, db)\}, \quad u_4 = babad.
\end{aligned}$$

Aide : pour (S_4, u_4), trouver des mots α, β tels que $bab^iad \to_{S_4}^* \alpha \cdot bab^{i+1}ad \cdot \beta$.
Le problème de correspondance de Post infinitaire (PCP-INF) est le problème suivant :

— Instance : une suite $I = ((u_1, v_1), \dots (u_n, v_n))$ de couples de mots finis sur un alphabet Σ

— Question : existe-t-il une suite infinie d'entiers $(i_k)_{k\in\mathbb{N}}$ telle que $u_{i_1}u_{i_2}\cdots u_{i_k}\cdots = v_{i_1}v_{i_2}\cdots v_{i_k}\cdots$?

Soit $S \subseteq A^*$ un système de réécriture de mots fini sur l'alphabet fini A et $u \in A^*$. Définissons $\Phi_1(S, u)$ en suivant l'idée de la partie IV, mais sans règle correspondant à v et sur un alphabet A de cardinalité finie quelconque ; c'est la suite de couples

$$(\#\#x, \#\#xu^x\#) \quad , ((ax, xa))_{a\in A}, (\#x, x\#), \quad ((\ell^x,\ {}^x r))_{(\ell,r)\in S}.$$

2- Montrer que, si $u \to_S^\infty$ alors le (PCP-INF) sur $\Phi_1(S, u)$ a une solution.
3- Est-il vrai que, si le (PCP-INF) sur $\Phi_1(S, u)$ a une solution alors $u \to_S^\infty$?

Considérons le cas où $A = A_0 \cup M$ (avec $A_0 \cap M = \emptyset$). Les lettres de M sont vues comme des lettres marquées. On appelle mot marqué tout mot $u \in A_0^* M A_0^*$ et système marqué toute partie $S \subseteq A_0^* M A_0^* \times A_0^* M A_0^*$ i.e. chaque membre gauche (resp. droit) de règle est un mot marqué.
Étant donné un système fini marqué S et un mot marqué u, on définit la suite de couples $\Phi_2(S, u)$ par :

$$(\#\#x, \#\#xu^x\#) \quad , ((ax, xa))_{a\in A_0}, (\#x, x\#), \quad ((\ell^x,\ {}^x r))_{(\ell,r)\in S}.$$

(on remarquera que les lettres de M n'apparaissent que dans $u^x, \ell^x,\ {}^x r$).
4- Montrer que, si S, u sont marqués et $u \to_S^\infty$, alors le (PCP-INF) sur $\Phi_2(S, u)$ a une solution.
5- Montrer que, si S, u sont marqués et le (PCP-INF) sur $\Phi_2(S, u)$ a une solution, alors $u \to_S^\infty$.
6- Donner une réduction du problème de l'arrêt des machines de Turing au problème (TERM-ST) pour des systèmes marqués et un mot de départ marqué.
7- Montrer que le problème (PCP-INF) est indécidable.