Option Informatique en Spé MP

Devoir surveillé du mardi 25 novembre 1997

Autour des idées d'Axel Thue :
systèmes de réécriture et régularités dans les mots infinis

Résumé

Au début du vingtième siècle, le mathématicien Axel THUE écrivit plusieurs articles fondant deux secteurs de l'informatique théorique : les *systèmes de réécriture*, qui interviennent par exemple dans les grammaires formelles, les méthodes de démonstration automatique, les logiciels de calcul formel ; et l'étude des *régularités* dans les mots infinis, laquelle trouve actuellement des applications en génétique moléculaire.

Les deux premières parties du problème, après quelques généralités, étudient la terminaison de plusieurs systèmes de réécriture très simples. On donne ensuite quelques notions sur les mots infinis, en particulier ceux qui sont obtenus par itération d'un morphisme.

La cinquième partie présente la notion de propriété inévitable, et caractérise les alphabets sur lesquels la propriété *être sans facteur carré* est évitable.

Enfin, la sixième partie montre que la suite des mouvements, dans la solution optimale du problème des tours de Hanoï, peut être décrite par l'itération d'un morphisme, et que cette suite ne contient aucun facteur carré ; ce dernier résultat a été établi en 1991 par Dan ASTOORIAN et Jim RANDALL, alors étudiants en licence à l'université de Waterlooo au Canada.

Veuillez rédiger chaque partie sur une copie séparée.

Table des matières

Notations, définitions, et mises en garde

► Dans tout l'énoncé, A est un alphabet fini. Le mot vide est noté ε. La longueur d'un mot u est notée $|u|$; $|u|_a$ désigne le nombre d'occurrences de la lettre a dans le mot u.

► u est *préfixe* de v s'il existe un mot w tel que $v = uw$; si $w \neq \varepsilon$, u est *préfixe propre* de v.

► La suite de FIBONACCI est définie par ses deux premiers termes $F_0 = 0$, $F_1 = 1$ et la relation $F_{n+2} = F_{n+1} + F_n$ pour tout $n \in \mathbb{N}$.

On peut admettre un résultat, à condition de le signaler clairement ; en tout état de cause, il est demandé de rédiger les questions dans l'ordre de l'énoncé. Les programmes devront être concis, et suffisamment documentés pour être compréhensibles. L'emploi de références est interdit. Les questions marquées $\star\star\star$ *sont, à mon sens, plus délicates.*

1 Quelques propriétés des relations

► On fixe un ensemble E. Une *relation* sur E est une partie $\mathcal{R}$ de $E \times E$; lorsque $(x, y) \in \mathcal{R}$, on note $x \to y$. On associe naturellement un graphe orienté à une relation $\mathcal{R}$: les sommets sont les éléments de E, les arcs sont les éléments de $\mathcal{R}$.

► Toutes les relations considérées dans la suite sont *acycliques* : si une suite $z_0, z_1, \ldots, z_n$ d'éléments de E vérifie $z_{i-1} \to z_i$ pour tout $i \in [\![1, n]\!]$ et $z_n \to z_0$, alors $z_0 = z_1 = \cdots = z_n$.

► On note $x \xrightarrow{*} y$ s'il existe une suite $z_0, z_1, \ldots, z_n$ d'éléments de E telle que $z_0 = x$, $z_n = y$ et $z_{i-1} \to z_i$ pour tout $i \in [\![1, n]\!]$. Clairement, la relation $\xrightarrow{*}$ est réflexive (prendre $n = 0$ dans la définition) et transitive.

► Une relation est *confluente* si $x \xrightarrow{*} y$ et $x \xrightarrow{*} z$ impliquent l'existence d'un t tel que $y \xrightarrow{*} t$ et $z \xrightarrow{*} t$. Une relation est *nœthérienne* s'il n'existe pas de suite infinie $(z_n)_{n\in\mathbb{N}}$ telle que $z_i \to z_{i+1}$ pour tout $i \in \mathbb{N}$.

► Un élément y de E est *irréductible* s'il n'existe aucun élément z distinct de y et tel que $y \to z$; y est un *réduit* de x si $x \xrightarrow{*} y$ et si y est irréductible.

Question 1 • Montrez que si $\mathcal{R}$ est nœthérienne, chaque élément de E possède au moins un réduit.

Question 2 • Exhibez une relation nœthérienne vérifiant la propriété suivante : pour tout $n \geqslant 1$, il existe un élément de E possédant au moins n réduits distincts.

Question 3 • Montrez que si $\mathcal{R}$ est confluente, chaque élément de E possède au plus un réduit.

Question 4 • Exhibez une relation $\mathcal{R}$ confluente mais non nœthérienne.

Question 5 • Soit $\mathcal{R}$ une relation nœthérienne. Montrez que $\mathcal{R}$ est confluente si et seulement si tout élément possède un et un seul réduit.

2 Systèmes de réécriture

► Dans cette partie, l'alphabet A contient au moins deux lettres a et b. Une *règle de réécriture* est un couple (u, v) de mots sur l'alphabet A ; la règle est notée $u \to v$.

► Un *système de réécriture* est un ensemble fini S de règles de réécriture. À un tel ensemble, on associe la relation sur A^* définie par $x \to y$ s'il existe une règle $u \to v$ appartenant à S et des mots s et t tels que $x = sut$ et $y = svt$; l'application de la règle consiste donc à remplacer une occurrence de u par v. On note que le choix $s = t = \varepsilon$ assure la cohérence des notations choisies.

Question 6 • Montrez que la relation $\xrightarrow{*}$ associée à un système de réécriture vérifie la propriété suivante : si $x \xrightarrow{*} y$, alors $sxt \xrightarrow{*} syt$ quels que soient les mots s et t.

Question 7 • Montrez que si l'on a $|y| < |x|$ pour chaque règle $x \to y$ d'un système de réécriture, alors la relation associée est nœthérienne.

► On considère dans les deux questions suivantes l'alphabet $A = \{a, b\}$ et la relation $\mathcal{R}$ associée à l'unique règle $ab \to bba$.

► Soit $u \in A^*$; on peut l'écrire $u = b^{n_0}ab^{n_1}a \ldots b^{n_{p-1}}ab^{n_p}$, où $n_0, n_1, \ldots, n_p$ sont des naturels, avec $p = |u|_a$. On note $\varphi(u) = \sum_{0\leqslant k\leqslant p} 3^k n_k$.

Question 8 • Soit v tel que $u \to v$. Montrez que $\varphi(v) < \varphi(u)$, en déduire que la relation $\mathcal{R}$ est nœthérienne.

Question 9 • En observant l'application ψ définie par $\psi(u) = \sum_{0 \leqslant k \leqslant p} 2^k n_k$, montrez que tout $u \in A^*$ possède un et un seul réduit, que l'on explicitera.

Question 10 • Dans cette question, A est un alphabet quelconque. Un système de réécriture S vérifie la propriété suivante : il existe des mots x, s et t tels que $x \overset{*}{\rightarrow} sxt$, avec $st \neq \varepsilon$. Montrez que la relation $\mathcal{R}$ associée à ce système n'est pas nœthérienne.

Question 11 • Que pensez-vous de la relation définie par la règle $ab \rightarrow bbaa$?

▶ On considère maintenant le système de réécriture sur l'alphabet $A = \{a, b, c\}$ constitué des règles $ac \rightarrow cb$, $bc \rightarrow ca$, $cc \rightarrow \varepsilon$ et $ab \rightarrow ba$.

Question 12 • Les mots *bacacbbcbc* et *acbaccaba* admettent-ils un réduit commun ?

Question 13 • La relation $\rightarrow$ est-elle nœthérienne ?

Question 14 • La relation $\rightarrow$ est-elle confluente ?

Question 15 • Rédigez en Caml une fonction de type `string -> string` qui calcule le réduit d'un mot pour la relation $\rightarrow$.

3 Mots infinis

▶ Un *mot infini* sur un alphabet A est une application $v : \mathbb{N}^* \mapsto A$. On note v_i au lieu de $v(i)$, et $v = v_1 v_2 \dots v_n \dots$ L'ensemble des mots infinis sur A est noté A^ω.

▶ Pour $1 \leqslant i \leqslant j$, on note $v[i..j] = v_i v_{i+1} \dots v_{j-1} v_j$. Un mot u est un *facteur* de v s'il existe des indices i et j tels que $u = v[i..j]$; si $i = 1$, on dit que u est un *préfixe* de v.

▶ Si $u \in A^*$, on note uv le mot infini défini par $(uv)_i = u_i$ si $1 \leqslant i \leqslant |u|$ et $(uv)_i = v_{i-|u|}$ si $i > |u|$. On a donc, de manière naturelle, $uv = u_1 \dots u_n v_1 v_2 \dots$

▶ Si $u \in A^*$ et $v \in A^+$, on note uv^ω le mot infini $uvvv\dots$ obtenu en concaténant une infinité d'exemplaires de v derrière le mot u. Par exemple, le mot $x = a(ba)^\omega$ est défini par $x_n = a$ si n est impair et $x_n = b$ si n est pair.

Question 16 • Soit L une partie infinie de A^+. Construisez un mot infini u sur A tel que chaque prefixe de u soit préfixe d'une infinité de mots de L.

Question 17 • À quel théorème la technique utilisée dans la question précédente vous fait-elle penser ?

Question 18 • Soit L une partie infinie de A^+. Prouvez l'existence d'un mot infini u sur A tel que tout facteur de u soit facteur d'une infinité de mots de L.

▶ Soit v un mot infini ; on note $\mathcal{L}_n(u)$ l'ensemble des facteurs de longueur n de u. La *complexité* de u est la suite de terme général $C_n(u)$, égal au cardinal de $\mathcal{L}_n(u)$.

Question 19 • Existe-t-il des mots infinis u vérifiant $C_n(u) = 1$ pour tout $n \in \mathbb{N}$?

Question 20 • Quelle est la complexité du mot $u = a(ba)^\omega$?

Question 21 • Exhibez un mot infini sur l'alphabet $A = \{a, b\}$, de complexité $C_n(u) = 2^n$.

Question 22 • Montrez que la complexité d'un mot infini est une suite croissante.

Question 23 • Justifiez la relation $C_{n_1+n_2}(u) \leqslant C_{n_1}(u) C_{n_2}(u)$.

▶ Un mot infini u est *ultimement périodique* s'il est de la forme xy^ω ; ceci revient à dire qu'il existe un indice n_1 et une période $p > 0$ tels que $u_{n+p} = u_n$ pour tout $n \geqslant n_1$ (prendre $n_1 = |x| + 1$ et $p = |y|$).

Question 24 • Montrez que la complexité d'un mot infini ultimement périodique est une suite stationnaire : il existe un indice n_0 tel que $C_{n+1}(u) = C_n(u)$ pour tout $n \geqslant n_0$.

Question 25 • Soit u un mot infini. On suppose qu'il existe un rang n tel que $C_{n+1}(u) = C_n(u)$. Montrez que ce mot infini est ultimement périodique.

Question 26 • Le professeur Jean-Benoît MALENCONTREUX affirme avoir découvert un mot infini u dont la complexité $C_n(u)$ est équivalente à $\ln n$ lorsque n tend vers l'infini. Qu'en pensez-vous ?

4 Mot infini défini par itération d'un morphisme

▶ Soit $(x_n)_{n\geqslant 1}$ une suite de mots (finis) sur A ; on suppose que x_n est préfixe propre de x_{n+1}. Clairement, il existe un et un seul mot infini y dont chaque x_n est préfixe ; nous dirons que la suite $(x_n)_{n\geqslant 1}$ *converge* vers y. **Attention** : ici, x_n ne désigne pas la n-ième lettre d'un mot x, mais le terme d'indice n de la suite $(x_n)_{n\geqslant 1}$.

Question 27 • Que peut-on dire de deux suites qui ont la même limite?

▶ Soit $\varphi : A \mapsto A^+$; φ se prolonge en un *morphisme* de A^* dans lui-même défini comme suit : $\varphi(\varepsilon) = \varepsilon$; et, si $u = u_1u_2\dots u_n$, alors $\varphi(u) = \varphi(u_1)\varphi(u_2)\dots\varphi(u_n)$. Clairement, $\varphi(uv) = \varphi(u)\varphi(v)$ quels que soient les mots (finis) u et v.

▶ Dans les deux questions suivantes, on suppose que a est une préfixe propre de $\varphi(a)$.

Question 28 • Montrez que la suite de terme général $x_n = \varphi^n(a)$ converge. On notera désormais $\varphi^\omega(a)$ le mot infini limite de cette suite.

▶ On définit une application Φ de A^ω dans lui-même comme suit : si u est le mot infini $u_1u_2\dots u_n\dots$, alors $\Phi(u)$ est le mot infini $\varphi(u_1)\varphi(u_2)\dots\varphi(u_n)\dots$

Question 29 • Montrez que $\varphi^\omega(a)$ est un point fixe de Φ.

▶ Dans la suite de cette partie, on fixe $A = \{a, b\}$ et on définit le morphisme φ par $\varphi(a) = ab$ et $\varphi(b) = abb$. On note $v_n = \varphi^n(a)$ et $v = \varphi^\omega(a)$.

Question 30 • Calculez v_2, v_3 et v_4.

Question 31 • Donnez des expressions *simples* de $x_n = |v_n|_a$ et $y_n = |v_n|_b$.

Question 32 • Montrez que ni aa, ni bbb ne sont facteurs de v

Question 33 • Énumérez les facteurs de v de longueur 3, puis ceux de longueur 4.

Question 34 • Rédigez en Caml une fonction :

```
facteurs : string -> int -> string list
```

qui dresse la liste de tous les facteurs de longueur donnée d'un mot donné.
Par exemple, `facteurs "acbbcbacba" 3` rendra (à l'ordre près) la liste suivante :
`["acb";"cbb";"bbc";"bcb";"cba";"bac"]`.
On pourra utiliser la fonction
`sub_string : string -> int -> int -> string`
`sub_string s début lgr` rend la chaîne de caractères de longueur `lgr`, contenant les caractères d'indice `début` à `début + lgr - 1` de la chaîne `s`. L'exception `Invalid_argument "sub_string"` est déclenchée si `début` et `lgr` ne désignent pas une sous-chaîne de `s`. Rappel : les chaînes de caractères sont indexées comme les vecteurs, donc de `0` à `string_length s - 1` pour la chaîne `s`.

Question 35 • En observant le mot v_2, montrez que tout facteur de v possède une infinité d'occurrences dans v.

5 Propriétés inévitables

▶ Soient A un alphabet, et $\mathcal{P}$ une assertion relative aux mots sur A : $\mathcal{P}$ est une application de A^* dans l'ensemble $\{\textsf{vrai}, \textsf{faux}\}$. On dit que $u \in A^*$ *vérifie* $\mathcal{P}$ lorsque $\mathcal{P}(u) = \textsf{vrai}$.

▶ $\mathcal{P}$ est *inévitable* si l'ensemble $\{u \mid \mathcal{P}(u) = \textsf{faux}\}$ des mots qui ne vérifient pas $\mathcal{P}$ est fini ; dans le cas contraire, $\mathcal{P}$ est *évitable*. $\mathcal{P}$ est *idéale* si xuy vérifie $\mathcal{P}$ dès que u vérifie $\mathcal{P}$.

Question 36 • Soit $\mathcal{P}$ une propriété idéale. Montrez que $\mathcal{P}$ est évitable ssi il existe un mot infini u dont aucun facteur ne vérifie $\mathcal{P}$.

▶ On s'intéresse à la propriété $\mathcal{P}$ définie comme suit : $\mathcal{P}(u) = \textsf{vrai}$ ssi u possède un facteur carré, *id est* il existe des mots r, s et t tels que $u = rs^2t$, avec $s \neq \varepsilon$.

Question 37 • Montrez que $\mathcal{P}$ est idéale, et qu'elle est inévitable si l'alphabet A ne contient pas plus de deux lettres.

▶ Dans la suite de cette partie, on fixe $A = \{a, b, c\}$ et on considère le morphisme φ défini par $\varphi(a) = abc$, $\varphi(b) = ac$ et $\varphi(c) = b$. On note $v = \varphi^\omega(a)$.

Question 38 • Montrez que φ est injectif.

Question 39 • Énumérez les facteurs de longueur 3 de v.

Question 40 • Montrez qu'aucun facteur de v ne vérifie la propriété $\mathcal{P}$.

► On a ainsi établi que la propriété $\mathcal{P}$ est inévitable si l'alphabet A contient au plus deux lettres, évitable dans le cas contraire.

Question 41 • La propriété $\mathcal{P}$ est-elle conservée par φ, *id est* : si le mot u vérifie $\mathcal{P}$, en est-il de même pour le mot $\varphi(u)$?

Question 42 • Rédigez en Caml une fonction de type `string -> bool` qui indique si un mot possède un facteur carré.

6 Les tours de Hanoï

► La solution optimale du problème des tours de Hanoï est bien connue : pour transférer une pile de n disques de l'aiguille α vers l'aiguille γ, on transfère les $n-1$ disques du dessus de la pile de l'aiguille α vers l'aiguille β, puis le plus grand disque de l'aiguille α vers l'aiguille γ, et enfin on transfère la pile de $n-1$ disques de l'aiguille β vers l'aiguille γ. Le coût de cette solution est $C_n = 2^n - 1$.

► On utilise un alphabet $\Sigma = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \overline{\mathbf{a}}, \overline{\mathbf{b}}, \overline{\mathbf{c}}\}$ pour coder les mouvements élémentaires (déplacements d'un disque), selon le tableau suivant :

1 vers 2	$\mathbf{a}$	2 vers 1	$\overline{\mathbf{a}}$
2 vers 3	$\mathbf{b}$	3 vers 2	$\overline{\mathbf{b}}$
3 vers 1	$\mathbf{c}$	1 vers 3	$\overline{\mathbf{c}}$

On note H_{2n} le mot codant la solution optimale pour le transfert de $2n$ disques de l'aiguille 1 vers l'aiguille 3 ; et H_{2n+1} le mot codant la solution optimale pour le transfert de $2n+1$ disques de l'aiguille 1 vers l'aiguille 2. Ainsi $H_0 = \varepsilon$, $H_1 = \mathbf{a}$ et $H_2 = \mathbf{a}\overline{\mathbf{c}}\mathbf{b}$.

Question 43 • Explicitez H_3.

► On note σ le morphisme défini par $\sigma(\mathbf{a}) = \mathbf{b}$, $\sigma(\mathbf{b}) = \mathbf{c}$, $\sigma(\mathbf{c}) = \mathbf{a}$, $\sigma(\overline{\mathbf{a}}) = \overline{\mathbf{b}}$, $\sigma(\overline{\mathbf{b}}) = \overline{\mathbf{c}}$ et $\sigma(\overline{\mathbf{c}}) = \overline{\mathbf{a}}$.

Question 44 • Montrez que si le mot x code le transfert d'une pile de disques de l'aiguille 1 vers l'aiguille 2, alors $\sigma(x)$ code le transfert de la même pile de l'aiguille 2 vers l'aiguille 3.

Question 45 • Établissez la relation $H_{2n+1} = H_{2n}\mathbf{a}\sigma^{-1}(H_{2n})$, ainsi qu'une relation analogue exprimant H_{2n+2} en fonction de H_{2n+1}.

Question 46 • Montrez que la suite de mots $(H_n)_{n\in\mathbb{N}}$ converge vers un mot infini $\mathcal{H}$, vérifiant les propriétés suivantes :

- $\mathcal{H}_{6k+1} = \mathbf{a}$ et $\mathcal{H}_{6k+2} \in \{\mathbf{c}, \overline{\mathbf{c}}\}$
- $\mathcal{H}_{6k+3} = \mathbf{b}$ et $\mathcal{H}_{6k+4} \in \{\mathbf{a}, \overline{\mathbf{a}}\}$
- $\mathcal{H}_{6k+5} = \mathbf{c}$ et $\mathcal{H}_{6k+6} \in \{\mathbf{b}, \overline{\mathbf{b}}\}$

► On définit un morphisme φ par les relations suivantes :

$\varphi(\mathbf{a}) = \mathbf{a}\overline{\mathbf{c}}$	$\varphi(\overline{\mathbf{a}}) = \mathbf{ac}$
$\varphi(\mathbf{b}) = \mathbf{c}\overline{\mathbf{b}}$	$\varphi(\overline{\mathbf{b}}) = \mathbf{cb}$
$\varphi(\mathbf{c}) = \mathbf{b}\overline{\mathbf{a}}$	$\varphi(\overline{\mathbf{c}}) = \mathbf{ba}$

Question 47 • Justifiez les relations $\varphi \circ \sigma = \sigma^{-1} \circ \varphi$ et $\varphi \circ \sigma^{-1} = \sigma \circ \varphi$.

Question 48 • Établissez les deux relations :

$$\varphi(H_{2n})\mathbf{a} = H_{2n+1} \qquad \text{et} \qquad \varphi(H_{2n+1})\mathbf{b} = H_{2n+2}$$

Question 49 • Prouvez alors que $\mathcal{H}$ est le point fixe du morphisme φ.

Question 50 • Montrez que $\mathcal{H}$ ne contient pas quatre lettres consécutives non surlignées.

Question 51 • En déduire que $\mathcal{H}$ n'a pas de facteur carré.

Question 52 • Le langage $\{H_n : n \in \mathbb{N}\}$ est-il rationnel ?

FIN