Lemme de l'étoile

Corrigé partiel de la feuille de travaux dirigés n^{o}6

16 mars 2009

1.

a) Supposons L_1 rationnel. On fixe n et on considère le mot $z = 1^p$ pour p un nombre premier tel que $|p| \geq n$. Quel que soit le découpage en trois mots u, v et w on a $|u| + |w| = p - k$ et $|v| = k$. On considère alors la longueur du mot uv^iw : $|uv^iw| = (p-k) + ki = p + k(i-1)$. Pour la valeur de $i = p+1$, on obtient le mot $uv^{p+1}w$ dont la longueur vaut $p(k+1)$ qui n'est plus un nombre premier, donc plus dans le langage ; une contradiction. On en déduit donc que L_1 n'est pas rationnel.

b) Supposons que L_2 rationnel. On fixe n et on considère le mot $z = 0^n10^{n+1} \in L_2$. Toute factorisation est de la forme $uv = 0^{n-k}$ avec $v = 0^l$ pour $l > 0$ et $w = 0^k10^{n+1}$ pour $k \geq 0$. Par application du lemme, le mot $0^{n-k-l}0^{il}0^k1^{n+1}$ devrait être dans le langage. Pour ce faire, l'égalité $n-k-l+il+k+1 = n+1$ doit toujours être vérifiée. Or, pour $i \neq 1$ le mot uv^iw n'est plus dans le langage ; une contradiction. On en déduit donc que L_2 n'est pas rationnel.

Autre preuve : Supposons L_2 rationnel. Alors, il existe un automate fini déterministe $\mathcal{A}$ à p états qui reconnaît L_2. En particulier, le mot $z = 0^{p+1}10^{p+2}$ doit être reconnu par $\mathcal{A}$. Puisque la longueur de ce mot dépasse le nombre d'états de l'automate, il existe nécessairement un état q sur lequel on passe plus d'une fois pour le préfixe 0^{p+1}. On a donc un chemin de longueur i allant de l'état initial de $\mathcal{A}$ vers l'état q, suivi d'un circuit de longueur $j \geq 1$ qui «boucle» sur l'état q puis un chemin de longueur k tel que $p+1 = i+j+k$ suivi d'un chemin de longueur $p+2$ qui arrive dans un état terminal. Comme z est reconnu, la lecture du mot où

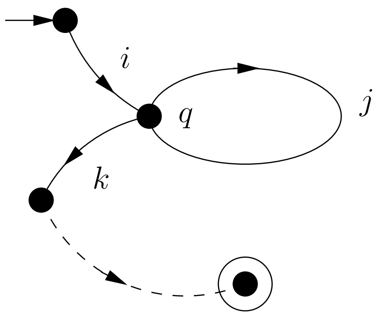

FIGURE 1 – Parcours sur l'automate $\mathcal{A}$ pour le mot z.

on a supprimé le circuit est également acceptante. Il s'ensuit que le mot $0^{i+j}10^{p+2}$ est reconnu. Or, puisque $i + j < p$, ce mot n'est pas dans le langage ; une contradiction. Le langage L_2 n'est donc pas rationnel.

c) Supposons L_3 rationnel. On fixe n et on considère le mot $z = a^kb^nc^n$ avec $0 < k < n$. Alors, pour la factorisation $u = \varepsilon, v = a^k$ et $w = b^nc^n$, le mot uv^iw devrait être dans le langage pour toute valeur de i. Ce qui n'est pas le cas si on choisit $i = 0$. Il en est de même pour la factorisation $u = a^{k-l}$, $v = a^lb^m$ et $w = b^{n-m}c^n$, puisque le facteur milieu itéré ne donne plus un mot du langage. Enfin, pour la factorisation où $u = a^k$, $v = b^m$ et $w = b^{n-m}c^n$, par itération du facteur milieu, on obtient un excédent de b. Dans tous les cas on arrive à une contradiction. On en déduit donc que L_3 n'est pas rationnel.

Autre preuve : Supposons L_3 rationnel. Soit alors $L = ab^\star c^\star$, un langage rationnel. Puisque L est rationnel et L_3 est supposé rationnel, $L \cap L_3$ doit être rationnel par les propriétés de clôture. Or $L \cap L_3 = ab^n c^n$ n'est pas rationnel. *On peut être tenté de s'arrêter ici en se disant que la concaténation d'un langage rationnel et d'un langage non rationnel n'est pas rationnelle. Ce raisonnement est faux. En effet, il suffit de prendre la concaténation de $(a+b)^\star$ avec $a^n b^n$ qui est rationnelle !* On pourrait le montrer en utilisant le lemme ou un raisonnement par automates. Dans les deux cas on obtient une contradiction. On en déduit que L_3 n'est pas rationnel.

2.
a) Soit $w \in L_1$. Alors $w = uvx$ avec $u = \varepsilon$, $v = aa$ et x le reste du mot w. Dans ce cas, $uv^+x \in L_1$ donc $uv^+x \in L$. En ce qui concerne uv^0x, le seul problème peut être $|x|_a = 0$, mais L_2 a été prévu exactement pour ce cas.
Le cas où w est un mot de L_2 est évident.
b) Etudions l'intersection $L \cap aab^\star c^\star$. Si L est rationnel, alors l'intersection l'est aussi. Il est facile de démontrer que cette intersection ($\{aab^jc^j | j \geq 0\}$) n'est pas rationnelle par le lemme de l'étoile.
c) Cet exercice nous permet d'apprécier le lemme de l'étoile à sa juste valeur. En effet, ce n'est pas un outil de preuve de rationnalité ; si un langage vérifie les conditions, cela ne prouve rien du tout.

3.
a) Supposons L_1 rationnel. On fixe n et on considère le mot $z = a^m$ pour m un carré tel que $m \geq n$. Soit k la longueur de u dans un découpage en trois mots u, v et w. On considère alors la longueur du mot uv^iw : $|uv^iw| = m + k(i-1)$. Ainsi, pour que le lemme de la pompe soit vérifié il faut que les nombres $m-k$, m, $m+k$, $m+2k$, ... soient tous des carrés, c.a.d. une suite arithmétique infinie composée que de carrés. Ceci est impossible, car la différence de deux carrés succésives croit, alors que la différences dans la suite arithmétique est constante. En effet, considérons les deux termes succéssives suivantes dans cette suite : $m+k^2$ et $m+k^2+k$. Si le premier est un carré, il est le carré de $t = \sqrt{m+k^2} \geq k$, mais alors la différence entre ce carré et le carré suivant $(t+1)^2$ est d'au moins $2t \geq 2k$. Donc le lemme ne peut pas être vérifié est ainsi L_1 n'est pas rationnel.
b) Le plus simple est d'utiliser les propriétés de clôture. En effet, si on fait la substitution $f(a) = a$, $f(b) = a$ et $f(c) = \varepsilon$ alors $f(L_2) = L_1$. Donc si L_2 serait rationnel alors L_1 serait aussi, ce qui n'est pas le cas.