

Contrôle Continu 2 : correction

1 Lemme d'Arden

Attention aux erreurs suivantes:

- On rappelle le lemme d'Arden:
L'équation $X = PX + R$, où $\epsilon \notin P$ (important), a pour solution $X = P^*R$ (et non pas $P^* + R$).
- ne pas oublier de justifier (au moins un peu) quand vous faites des simplifications sur les expressions régulières.
- sauf exceptions, il n'y a pas commutation entre les lettres: $c(ac)$ n'est pas ac^2 .
- pour un langage $L + L = L$. Ainsi $ac + ac = ac$ et pas $2ac$.
- ne pas oublier les ϵ .

Pour une résolution facile, il est conseillé de faire le maximum de substitutions avant d'appliquer le lemme d'Arden (parce qu'il fait grossir énormément les expressions).

On peut utiliser les remarques suivantes:

- si $X_1 = X_2 + Y$ (où X_1 , X_2 et Y sont des ensembles de mots), alors $X_1 + X_2 = X_1$, puisque $X_2 \subset X_1$
- si deux des X_i que vous avez à calculer ont EXACTEMENT la même expression, ils sont égaux

Pour l'automate proposé:

$$\left\{ \begin{array}{l} X_0 = aX_1 \\ X_1 = cX_2 + dX_3 \\ X_2 = aX_1 + aX_4 + \epsilon \\ X_3 = aX_1 + aX_4 + \epsilon \\ X_4 = dX_2 + cX_3 + cX_5 \\ X_5 = \epsilon \end{array} \right.$$

Première remarque: $X_2 = X_3$

Alors $X_1 = (c + d)X_2$ et $X_4 = (c + d)X_2 + cX_5 = (c + d)X_2 + c$. Donc $X_4 = X_1 + c$.

En remplaçant dans l'expression de X_2 (ou X_3): $X_2 = a(X_1 + X_4) + \epsilon$ et par la remarque du début, $X_2 = aX_4 + \epsilon = aX_1 + ac + \epsilon$.

D'où $X_2 = (ac + ad)X_2 + c + \epsilon$; comme $\epsilon \notin (ac + ad)$, on peut appliquer le lemme d'Arden $\Rightarrow X_2 = (ac + ad)^*(ac + \epsilon)$.

En combinant avec la première équation: $X_0 = aX_1 = a(c + d)(ac + ad)^*(ac + \epsilon)$. CQFD.

On peut simplifier un peu l'expression: $X_0 = (ac + ad)(ac + ad)^*(ac + \epsilon)$. On a le droit de commuter un langage avec lui-même (ie $L.L^2 = L^2.L$, et en particulier ici $L.L^* = L^*.L$, mais les cas de commutation sont rares et c'est donc à éviter), d'où $X_0 = (ac + ad)^*(ac + ad)(ac + \epsilon)$. On peut enfin faire la remarque que $(ac + ad)^*(ac + ad)ac \subset (ac + ad)^*(ac + ad)$ et obtenir $X_0 = (ac + ad)^*(ac + ad)$.

2 Lemme de l'étoile

Quand vous devez démontrer qu'un langage est reconnaissable, vous pouvez oublier le lemme de l'étoile. Ne vous sentez pas obligé de l'utiliser dès que l'énoncé contient le mot "étoile" vu qu'il ne sert à rien pour les langages reconnaissables. Pour démontrer qu'un langage est reconnaissable: soit vous donnez un automate qui le reconnaît (en le justifiant), soit une expression régulière qui le décrit (en la justifiant), soit vous obtenez le langage par opérations ensemblistes sur des langages reconnaissables.

1. Le langage $L_1 = \{u \in \{a, b\}^* \mid |u|_{ab} - |u|_{ba} \leq 0\}$.

Réponses aux petites questions annexes:

- entre deux occurrences consécutives de ba il y a une et une seule occurrence de ab . ATTENTION: ce n'est vrai que dans l'alphabet $\{a, b\}$; par exemple $bacbac$ ne contient pas de ab entre les 2 occurrences de ba . Vous pouvez essayer de le démontrer (proprement) si vous voulez
- en conséquence de la première remarque, $|u|_{ab} - |u|_{ba}$ vaut -1 , 0 ou 1 .

Comme $|u|_{ab} - |u|_{ba} \in \{-1, 0, 1\}$, on peut prévoir qu'un automate sera facile à construire. Le seul détail à prendre en compte est que le comptage des occurrences de ab et ba ne peut commencer qu'à la deuxième lettre (0 ou 1 lettre ne suffit pas à créer une occurrence). Il faut donc séparer le cas où la première lettre du mot est un a de celui où c'est un b . Au final, on trouve un automate à $2 \times 3 = 6$ états, chaque état correspondant à une paire (première lettre / $|u|_{ab} - |u|_{ba}$).

Une expression régulière pour ce langage est $\epsilon + a(a^*bb^*a)^* + b(b^*aa^*b)^*$

On peut remarquer que $L = \{u \in \{a, b, c\}^* \mid |u|_{ab} - |u|_{ba} \leq 0\}$ n'est plus reconnaissable, justement parce qu'on a plus $|u|_{ab} - |u|_{ba} \in \{-1, 0, 1\}$ (dur à prouver).

2. $L_2 = \{u \in \{a, b\}^* \mid |u|_{aa} - |u|_{ba} \leq 0\}$

Réponses aux questions annexes:

- $\forall k \geq 2$, $|a^k|_{aa} = k - 1$. Ainsi aaa contient 2 occurrences différentes (mais pas disjointes) de aa .
- $\forall k$, $|(ba)^k|_{ba} = k$ parce que 2 occurrences ne peuvent pas se chevaucher.

De la même manière que $L = \{a^p b^q \mid p \leq q\}$, on va utiliser le lemme de l'étoile pour augmenter le nombre d'occurrence de aa dans un mot donné, et obtenir une contradiction.

Supposons que L_2 soit reconnaissable; alors \exists un A.F.D. \mathcal{A} le reconnaissant: $L_2 = L(\mathcal{A})$. Notons $n = |\mathcal{A}|$ le nombre d'états de \mathcal{A} .

Pour $k > \max(\{n; 2\})$, le mot $u = a^k (ba)^k$ est tel que $|u|_{aa} = k - 1$ et $|u|_{ba} = k$, donc $u \in L_2$. Attention, le mot inverse $u' = (ba)^k a^k$ est tel que $|u'|_{aa} = k$ (et non $k - 1$) à cause du dernier a du morceau $(ba)^k$.

On va utiliser le facteur $v = a^k$ de u pour le lemme de l'étoile. Comme on a choisi $k > n$, on est dans les conditions d'application du lemme de l'étoile, et donc: $\exists v_1, v_2 \neq \epsilon, v_3$ tels que $v = v_1 v_2 v_3$ et $\forall i, w_i = v_1 v_2^i v_3 (ba)^k \in L_2$.

Comme v_1 , v_2 et v_3 sont des facteurs de $v = a^k$, $v_1 = a^{|v_1|}$, $v_2 = a^{|v_2|}$ et $v_3 = a^{|v_3|}$. De plus $|v_1| + |v_2| + |v_3| = |v| = k$. Donc $w_i = a^{k+(i-1)|v_2|}(ba)^k$. Ce qui conduit à : $|w_i|_{aa} = k - 1 + (i - 1)|v_2|$ et $|w_i|_{ba} = k$. D'après le lemme de l'étoile, $|v_2| > 0$, donc pour $i > \frac{1}{|v_2|} + 1$, on aura $|w_i|_{aa} - |w_i|_{ba} > 0$ et donc $w_i \notin L_2$. Ce qui contredit le $w_i \in L_2$ donné par notre utilisation du lemme de l'étoile. CQFD.