

Option Informatique en Spé MP et MP*

TD : codes, algorithme de Sardinas et Patterson : le corrigé

Quelques résultats simples sur les codes

Question 1 • L'ensemble des mots de la forme 0^n1 , où n décrit \mathbb{N} , répond à la question.

Question 2 • La réponse est négative comme en témoigne le mot $0100101 = (0100)(101) = (01)(001)(01)$.

Question 3 • Soit u appartenant à L^+ . Il s'agit de montrer que u possède exactement une factorisation. Nous allons raisonner par récurrence sur $|u|$. Le résultat est clair si $|u| < 5$. Supposons-le acquis pour tout mot de longueur n , où $n \geq 4$. Soit u de longueur $n + 1$; distinguons quatre cas de figure selon le préfixe u_1u_2 de longueur 2 de u . Le cas $u_1u_2 = 11$ ne peut se présenter. Si $u_1u_2 = 10$, alors u commence par 101 et le mot v défini par $u = 101v$ appartient à L^* ; de par l'hypothèse de récurrence, le résultat est clair. Si $u_1u_2 = 00$, alors u commence par 001 : même conclusion en faisant intervenir le mot v défini par $u = 001v$. Si $u_1u_2 = 01$, alors u commence par 01101 , 0101 , 01000 ou 01001 ; on conclut en faisant intervenir le mot v défini respectivement par $u = 101v$, $u = 01v$, $u = 01000v$ ou $u = 01001v$.

Question 4 • Considérons une égalité non banale de la forme $u_1u_2 \dots u_n = v_1v_2 \dots v_p$, où les u_i et les v_j sont tous dans X . Nous pouvons supposer $n + p$ minimale. Si u_1 est plus court que v_1 , alors u_1 est préfixe de v_1 : ceci est exclu; par raison de symétrie, v_1 ne peut être plus long que u_1 . Donc $|u_1| = |v_1|$, puis $u_1 = v_1$ ce qui nous ramène à l'égalité $u_2 \dots u_n = v_2 \dots v_p$, qui est non banale, mais ne fait intervenir que $n + p - 2$ mots de L : d'où la contradiction.

Question 5 • La réciproque est fautive: considérez par exemple le code banal $\{0, 1\}$.

Question 6 • Sens direct: soient s et t appartenant à Δ^* , tels que $\Phi(s) = \Phi(t)$. Notons $n = |s|$, $p = |t|$, $s = s_1s_2 \dots s_n$ et $t = t_1t_2 \dots t_p$. Comme $\varphi(x) \neq \varepsilon$ pour toute lettre $x \in \Delta$, nous avons soit $n = p = 0$ (cas banal), soit $n > 0$ et $p > 0$. $\Phi(s) = \varphi(s_1)\varphi(s_2) \dots \varphi(s_n)$ est égal à $\Phi(t) = \varphi(t_1)\varphi(t_2) \dots \varphi(t_p)$. Mais les $\varphi(s_i)$ et les $\varphi(t_j)$ sont des mots de X , qui est un code; donc $n = p$, puis $s_i = t_i$ pour tout $i \in \llbracket 1, n \rrbracket$. Finalement, Φ est injectif.

• Réciproque: soit $w \in \Sigma^*$ possédant deux factorisations: $w = u_1u_2 \dots u_n = v_1v_2 \dots v_p$, les u_i et v_j étant dans X . Notons $s_i = \varphi^{-1}(u_i)$ et $t_j = \varphi^{-1}(v_j)$. Alors $\Phi(s_1s_2 \dots s_n) = \Phi(t_1t_2 \dots t_p)$; comme Φ est injectif, ceci implique $n = p$, puis $s_i = t_i$, donc $u_i = v_i$ pour tout $i \in \llbracket 1, n \rrbracket$. Nous en déduisons que les deux factorisations sont identiques, donc X est un code.

L'algorithme de Sardinas et Patterson

Question 7 • Nous avons $P = X \cup \{0, 1, 00, 10, 010\}$. La figure 1 représente le graphe associé; remarquons que tous les arcs sont croisés.

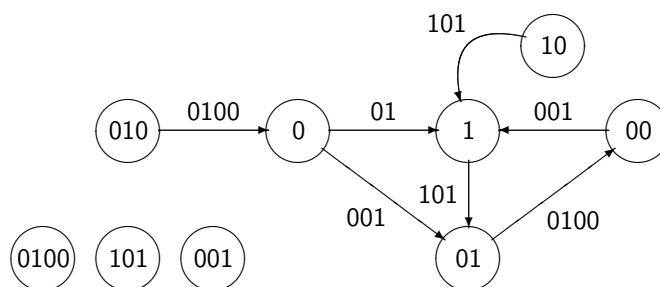


Figure 1: le graphe associé au langage M de la question 2

Question 8 • Si le chemin menant de x à u est de longueur 1, il se réduit à l'arc (x, u) . Si cet arc est croisé, alors $xu \in X$, donc $y = x$ et $z = u$ conviennent. Sinon, l'arc est direct : il existe $x' \in X$ tel que $xx' = u$; cette fois, $y = \varepsilon$ et $z = xx'$ conviennent.

• Soit maintenant $n \geq 1$; supposons le résultat acquis pour tout chemin de longueur inférieure ou égale à n . Soient $x \in X$ et $u \in P - X$ tels qu'il existe un chemin de longueur n menant de x à u . Notons (v, u) l'arc par lequel se termine ce chemin. Si $v \in X$, il suffit d'appliquer le résultat établi pour $n = 1$. Sinon, $v \in P - X$. L'hypothèse de récurrence appliquée au chemin (de longueur n) menant de x à v assure l'existence de y' et z' appartenant à X^* tels que $y'v = z'$. Si (v, u) est croisé, alors $vu \in X$, donc $y'vu = z'u$, si bien que $y = z'$ et $z = y'vu$ conviennent. Sinon, il existe $t \in X$ tel que $vt = u$; alors $y'u = y'vt = z't$, si bien que $y = y'$ et $z = z't$ conviennent. Ceci établit l'assertion au rang $n + 1$ et termine la preuve.

Question 9 • Nous allons raisonner par récurrence sur la longueur commune n des deux membres de l'égalité $yu = z$. Remarquons que z se termine par un mot w appartenant à X ; que ce mot ne peut être égal à u , puisque ce dernier appartient à $P - X$; et que ni u , ni w , ne peuvent être le mot vide ; donc $|u| \neq |w|$, puis $|u| \geq 2$ ou $|w| \geq 2$ et par suite $|z| \geq 2$.

• Si $|z| = 2$, alors $z = w$, puis $|y| = 1$ et donc $y \in X$. Par suite, (y, u) est un arc croisé et c'est un chemin qui mène de y à u .

• Soit $n \geq 2$. Supposons le résultat acquis pour toute égalité entre mots de longueur au plus n . Observons une égalité entre mots de longueur $n + 1$, en reprenant les notations précédentes. Notons t le mot de X^* défini par $z = tw$; nous aurons $yu = tw$. Comme $z \in X$ et $u \in P - X$, on a certainement $u \neq w$, et donc $|u| \neq |w|$, si bien que l'un des deux mots u et w est suffixe propre de l'autre, ce qui nous amène à distinguer deux cas de figure.

• Si $|u| < |w|$, alors u est suffixe propre de w : il existe $v \in P$ tel que $vu = w$; mais du coup (v, u) est un arc croisé de G . Si $v \in X$, alors l'arc (u, v) est un chemin qui répond à la question. Sinon, $v \in P - X$, et $tv = y$; comme t et y appartiennent à X^* et v à $P - X$, nous pouvons appliquer l'hypothèse de récurrence : il existe un sommet $x \in X$ et un chemin menant de x à v . Avec l'arc menant de v à u , nous obtenons un chemin menant de x à u .

• Si $|u| > |w|$, alors w est suffixe propre de u : cette fois, $u = vw$ avec $v \in P$, et (v, u) est un arc direct de G . Si $v \in X$, alors l'arc (u, v) est un chemin qui répond à la question. Sinon, $v \in P - X$, et $yv = t$; mais $|t| < |z|$, et on conclut à nouveau par récurrence.

Question 10 • Considérons un chemin de longueur $n \geq 1$ menant de s à t , avec s et t appartenant à X . Notons (u, t) le dernier arc de ce chemin ; comme $t \in X$, cet arc est croisé, donc le mot $v = ut$ appartient à X , et $v \neq t$ puisque $u \in P$. Si $u \in X$, alors $v \in X \cap X^2$, donc X n'est pas un code. Sinon, $u \in P - X$; d'après le lemme, il existe y et z appartenant à X^* tels que $yu = z$. Alors $yv = yut = zt$; or t et v sont dans X , avec $t \neq v$. Donc X n'est pas un code.

Question 11 • X n'est pas un code, donc il existe des mots y' et z' appartenant à X^* et y et z appartenant à X tels que $y'y = z'z$ et $y \neq z$; nous pouvons supposer $|z| > |y|$, si bien que y est suffixe de z : il existe u tel que $z = uy$; alors $u \in P$, et l'arc (u, y) est croisé. Si $u \in X$, alors le chemin réduit à l'arc (u, y) est le chemin cherché. Sinon, $y'y = z'z = z'uy$, donc $y' = z'u$; comme y' et z' appartiennent à X^* , le lemme assure l'existence de $s \in X$ tel qu'il existe un chemin menant de s à u ; en le prolongeant avec l'arc (u, y) , nous obtenons un chemin menant de s à y .

Question 12 • On commence par construire P , puis le graphe G associé. La recherche d'un chemin menant d'un sommet $s \in X$ à un sommet $t \in X$ (pas nécessairement distinct de s) se fait avec l'algorithme de parcours en largeur de G .

► Notons que dans le graphe présenté à la figure 1, il existe un chemin menant du sommet $01 \in X$ à lui-même ; ce qui nous donne une deuxième preuve du fait que le langage M n'est pas un code.

Question 13 • Notons $\lambda = \sum_{x \in X} |x|$. Énumérer P a un coût $\mathcal{O}(\lambda)$. Comme $|P| \leq \lambda$, le coût de la construction

du graphe G est certainement un $\mathcal{O}(\lambda^3)$. Enfin, pour chaque $x \in X$, le parcours en largeur de G en partant de X a un coût $\mathcal{O}(\lambda^2)$. Le coût total est donc polynomial en λ .

FIN