Option Informatique en Spé MP et MP*
Étude d'un mot infini : le corrigé

▶ Quelques remarques préliminaires :

1. si un facteur v de u est récurrent, alors tout facteur de v est lui-même récurrent ;
2. inversement, un mot v ayant un facteur non récurrent de u parmi ses facteurs ne peut être un facteur récurrent de u ;
3. en particulier, tout mot v ayant un mot de L parmi ses facteurs possède au plus une occurrence dans u ;
4. pour chaque $k \geqslant 1$, le mot $\mathbf{0}\,\mathbf{1}^k\,\mathbf{0}$ est facteur de u ; mieux, il possède exactement une occurrence dans u.

Question 1 • La k-ième occurrence de $\mathbf{0}$ est précédée du mot $\mathbf{0\,1\,0\,1\,1}\ldots\mathbf{0\,1}^{k-1}$, dont la longueur est :

$$\sum_{1\leqslant i<k} |\,\mathbf{0\,1}^i\,| = \sum_{1\leqslant i<k} (i+1) = k-1+\sum_{1\leqslant i<k} i = k-1+\frac{k(k-1)}{2} = \frac{k(k+1)}{2}-1$$

Donc $\boxed{P_k = \dfrac{k(k+1)}{2}}$.

Question 2 • Pour $k \geqslant 1$, le mot $\mathbf{0\,1}^k\,\mathbf{0}$ n'a qu'une occurrence, en position $\dfrac{k(k+1)}{2}$; $\mathbf{0\,1}^k$ a une infinité d'occurrences, aux positions P_j avec $j \geqslant k$; et $\mathbf{1}^k\,\mathbf{0}$ a lui aussi une infinité d'occurrences, aux positions $P_j - k$ avec $j > k$. Donc $\mathbf{0\,1}^k\,\mathbf{0}$ est un élément minimal de L.

• D'après la remarque 3, tout facteur v de u admettant le facteur $\mathbf{0\,1}^k\,\mathbf{0}$ appartient à L. En particulier, si v contient au moins deux occurrences de $\mathbf{0}$, il possède un facteur de la forme $\mathbf{0\,1}^k\,\mathbf{0}$ (puisque ces deux occurrences ne peuvent être consécutives), donc est un mot de L. Considérons maintenant un facteur v de u contenant au plus une occurrence de $\mathbf{0}$. Deux cas peuvent se présenter :

- $\mathbf{0}$ ne possède aucune occurrence dans v, donc $v = \mathbf{1}^p$ avec $p \geqslant 1$; ce mot est facteur de chaque $\mathbf{0\,1}^k\,\mathbf{0}$ pour $k \geqslant p$;
- $\mathbf{0}$ possède exactement une occurrence dans v, donc $v = \mathbf{1}^p\,\mathbf{0\,1}^q$ avec $p \geqslant 0$ et $q \geqslant 0$; ce mot est facteur de $\mathbf{0\,1}^k\,\mathbf{0\,1}^{k+1}\,\mathbf{0}$ pour chaque $k \geqslant \max(p, q-1)$.

Dans chaque cas, v est récurrent, donc n'appartient pas à L.

• Concluons : $\boxed{\text{les mots minimaux de } L \text{ sont les mots de la forme } \mathbf{0\,1}^k\,\mathbf{0}, \text{ avec } k \geqslant 1}$.

Question 3 • Ceci a été établi implicitement à la question précédente : les facteurs de u n'appartenant pas à L sont de la forme $\mathbf{1}^p$ avec $p \geqslant 1$ ou $\mathbf{1}^p\,\mathbf{0\,1}^q$ avec $p \geqslant 0$ et $q \geqslant 0$, et sont récurrents.

Question 4 • Les facteurs de v qui appartiennent à L possèdent exactement une occurrence ; les facteurs de v qui n'appartiennent pas à L sont récurrents. Concluons : les facteurs récurrents de u sont les mots de la forme $\mathbf{1}^p$ avec $p \geqslant 1$ ou $\mathbf{1}^p\,\mathbf{0\,1}^q$ avec $p \geqslant 0$ et $q \geqslant 0$.

Question 5 • $\mathbf{0\,0}$ n'est pas facteur de u. $\mathbf{0\,1}$ est facteur spécial de u, car $\mathbf{0\,1\,0}$ et $\mathbf{0\,1\,1}$ sont facteurs de u ; même raisonnement et même conclusion pour $\mathbf{1\,1}$. Enfin, $\mathbf{1\,0\,1}$ est facteur de u, mais $\mathbf{1\,0\,0}$ ne l'est pas, donc $\mathbf{1\,0}$ est facteur ordinaire de u.

• De la même façon, nous montrons que $\mathbf{0\,0\,0}$, $\mathbf{0\,0\,1}$ et $\mathbf{1\,0\,0}$ ne sont pas facteurs de u ; $\mathbf{0\,1\,1}$ et $\mathbf{1\,1\,1}$ sont facteurs spéciaux ; et $\mathbf{0\,1\,0}$, $\mathbf{1\,0\,1}$ et $\mathbf{1\,1\,0}$ sont facteurs ordinaires.

Question 6 • Notons α_n (resp. β_n) le nombre de facteurs ordinaires (resp. spéciaux) de longueur n ; ainsi, $C_u(n) = \alpha_n + \beta_n$. Un facteur ordinaire ne possède qu'un prolongement à droite, un facteur spécial en possède deux. Donc :

$$C_u(n+1) = \alpha_n + 2\beta_n = \alpha_n + \beta_n + \beta_n = C_u(n) + \beta_n$$

Ainsi, $C_u(n+1) - C_u(n)$ est bien égal à β_n.

Question 7 • Un facteur spécial de u contient au plus une occurrence de $\mathbf{0}$. Les facteurs spéciaux de longueur $n \geqslant 1$ de u sont donc de deux sortes :

- le mot $\mathbf{1}^n$: en effet, $\mathbf{1}^n\,\mathbf{0}$ apparaît dans $\mathbf{0}\,\mathbf{1}^n\,\mathbf{0}$, et $\mathbf{1}^n\,\mathbf{1} = \mathbf{1}^{n+1}$ apparaît dans $\mathbf{0}\,\mathbf{1}^{n+1}\,\mathbf{0}$;
- les mots de la forme $\mathbf{1}^j\,\mathbf{0}\,\mathbf{1}^{n-1-j}$ où j vérifie $0 \leqslant j \leqslant n-2-j$; en effet, $\mathbf{1}^j\,\mathbf{0}\,\mathbf{1}^{n-1-j}\,\mathbf{0}$ apparaît dans $\mathbf{0}\,\mathbf{1}^{n-2-j}\,\mathbf{0}\,\mathbf{1}^{n-1-j}\,\mathbf{0}$ et $\mathbf{1}^j\,\mathbf{0}\,\mathbf{1}^{n-1-j}\,\mathbf{1} = \mathbf{1}^j\,\mathbf{0}\,\mathbf{1}^{n-j}$ apparaît dans $\mathbf{0}\,\mathbf{1}^{n-1-j}\,\mathbf{0}\,\mathbf{1}^{n-j}\,\mathbf{0}$.

Si $j \geqslant n-1-j$, alors $\mathbf{1}^j\,\mathbf{0}\,\mathbf{1}^{n-1-j}$ est un facteur ordinaire : en effet, il n'apparaît que dans les facteurs de u de la forme $\mathbf{0}\,\mathbf{1}^k\,\mathbf{0}\,\mathbf{1}^{k+1}\,\mathbf{0}$ avec $k \geqslant \max(j, n-1-j) = j$; et il est alors suivi d'un $\mathbf{1}$.

Il nous reste à déterminer combien il existe d'indices j tels que $0 \leqslant j \leqslant n-2-j$: cette condition équivaut à $0 \leqslant 2j \leqslant n-2$, soit $0 \leqslant j \leqslant n/2-1$. Il existe donc $\lfloor n/2 \rfloor$ facteurs spéciaux du deuxième type. Au total, il y en a donc $1 + \lfloor n/2 \rfloor$. Remarquons que cette formule reste valable pour $n = 0$, puisque ε est spécial.

Question 8 • Il suffit d'appliquer les résultats des questions 6 et 7 et de faire un télescopage : Donc :

$$C_u(n) = C_u(0) + \sum_{0 \leqslant k < n} \big(C_{k+1}(u) - C_k(u)\big) = 1 + \sum_{0 \leqslant k < n} \left(1 + \left\lfloor \frac{k}{2} \right\rfloor\right) = n + 1 + \sum_{0 \leqslant k < n} \left\lfloor \frac{k}{2} \right\rfloor$$

Pour $n = 2p$, il vient, en regroupant les termes par 2 :

$$C_u(2p) = 2p + 1 + \sum_{0 \leqslant k < p} \left(\left\lfloor \frac{2k}{2} \right\rfloor + \left\lfloor \frac{2k+1}{2} \right\rfloor \right) = 2p + 1 + 2 \sum_{0 \leqslant k < p} k = 2p + 1 + 2\frac{(p-1)p}{2} = \boxed{p^2 + p + 1}$$

$C_u(2p+1) - C_u(2p)$ est le nombre de facteurs spéciaux de longueur $2p$, soit $p+1$ d'après la question 7. Donc $\boxed{C_u(2p+1) = p^2 + 2p + 1}$.

En revenant à n, nous obtenons $C_u(n) = \dfrac{n^2 + 2n + 4}{4}$ lorsque n est pair et $C_u(n) = \dfrac{n^2 + 2n + 5}{4}$ lorsque n est impair. Un peu de flair nous donne la formule $C_u(n) = \left\lfloor \dfrac{n^2 + 2n + 5}{4} \right\rfloor$.

Question 9 • Il existe des facteurs arbitrairement longs (de la forme $\mathbf{1}^k$) ne contenant pas le facteur $\mathbf{0}$, donc ce dernier (et tout facteur le contenant) n'est pas uniformément récurrent.

• Il nous reste les facteurs de la forme $\mathbf{1}^k$, dont on sait déjà qu'ils sont récurrents ; montrons qu'ils le sont uniformément. Notons Q_k la position dans u de la première occurrence de $\mathbf{1}^k$; tout facteur de longueur $2k$ situé à partir de cette position est de l'une des deux formes $\mathbf{1}^k\,\mathbf{0}\,\mathbf{1}^{k-1}$ ou $\mathbf{1}^{k-1}\,\mathbf{0}\,\mathbf{1}^k$, et admet donc le facteur $\mathbf{1}^k$. Nous en déduisons que tout facteur de u de longueur au moins $\max(2k, Q_k + k - 1)$ admet $\mathbf{1}^k$ comme facteur. Conclusion : $\mathbf{1}^k$ est uniformément récurrent.

Question 10 • La longueur de plus court préfixe de u contenant $\mathbf{1}^k$ est

$$\lambda_k = Q_k + k - 1 = P_k + k = \frac{k(k+1)}{2} + k = \frac{k(k+3)}{2}$$

Remarquons que $k(k+3) = k^2 + 3k \geqslant 4k$, donc $\lambda_k \geqslant 2k$. Finalement $\boxed{\rho(\mathbf{1}^k) = \dfrac{k(k+3)}{2}}$. Le tableau ci-dessous donne les premières valeurs de $\rho(\mathbf{1}^k)$:

k	1	2	3	4	5
$\rho(\mathbf{1}^k)$	2	5	9	14	20

Notes diverses

► L'étude des mots infinis ne date pas d'hier : ainsi, au début du vingtième siècle, Axel Thue a écrit deux articles importants, où il prouve en particulier l'existence d'un mot infini sans facteur carré sur un alphabet à trois lettres. Les mots *sturmiens* jouent un rôle central : ce sont les mots non ultimement périodiques de complexité minimale. Le point le plus récent sur ce sujet est le chapitre 2 du livre *Algebraic Combinatorics on Words*, de M. Lothaire.

► Les notions de complexité, récurrence, récurrence uniforme s'étendent aux *pavages*, qui sont des mots infinis bidimensionnels. Voir par exemple l'article *Tilings and quasiperiodicity* de Bruno Durand (École Normale Supérieure de Lyon), paru dans le volume 1256 des *Lecture Notes in Computer Science.* En principe, il se trouve sur le Web !

FIN

`[motinfinicor]` *Version du 24 février 2008*