Mots sans carré et mots sans cube (2019)

Sujet
*

Remarque : initialement, ce sujet a été construit pour être fait en 1h avec aide d'un examinateur. Il a été adapté pour le format actuel de l'épreuve en 30 minutes de préparation + 30 minutes de passage.

Dans ce sujet, les lettres d'un mot de longueur n sur un alphabet Σ seront notées $w[0], \dots, w[n-1]$. Un mot non-vide u est un *carré* s'il existe $v \in \Sigma^*$ tel que $u = v \cdot v$, et c'est un *cube* s'il existe $v \in \Sigma^*$ tel que $u = v \cdot v \cdot v$, où $\cdot$ dénote la concaténation.
Un mot $w \in \Sigma^*$ est *sans carré* si aucun de ses facteurs n'est un carré, et les mots sans cube sont définis de manière analogue. L'objet de ce sujet est de construire des mots sans carré et sans cube de longueur arbitrairement grande.

Question 1

Si l'alphabet Σ a une seule lettre, quelle est la plus grande longueur possible pour un mot sans carré ? pour un mot sans cube ?

Solution.

Si $\Sigma = \{a\}$, le seul mot a deux lettres est aa (qui est un carré) : donc tout mot de longueur au moins 2 admet aa comme facteur, donc contient un carré. La longueur maximale est 1 (avec le mot a).
Un raisonnement analogue donne une longueur maximale de 2 pour un mot sans cube (avec le mot aa).

Question 2

Si l'alphabet Σ a deux lettres, quelle est la plus grande longueur possible pour un mot sans carré ?

Solution.

Construisons un mot m sans carré sur l'alphabet $\{a, b\}$. Quitte à prendre le cas symétrique, on suppose que $m[0] = a$. Remarquons alors que $m[1] \neq a$: sinon, le préfixe serait aa, donc m contiendrait un carré. Donc $m[1] = b$. De même, $m[2] \neq b$, sinon $m[1]m[2] = bb$. Donc $m[2] = b$ (et notre mot serait aba, qui est sans carré). Supposons absurdement que $|m| \geqslant 4$: il faudrait alors choisir $m[3]$. Par le même raisonnement qu'avant, $m[3] \neq a$; donc $m[3] = b$. Mais alors $m = abab$, qui est un carré : absurde !
Donc la longueur maximale est 3.

Jusqu'à la question 6, on considère l'alphabet $\Sigma = \{a, b\}$. On définit la fonction $h : \Sigma \to \Sigma^*$ par $h(a) = ab$ et $h(b) = ba$, et on l'étend à une fonction de Σ^* dans Σ^* de la façon suivante : pour $w = w[0] \cdots w[n-1]$ un mot de Σ^*, on note $h(w) := h(w[0]) \cdots h(w[n-1])$. En particulier, pour ε le mot vide, on a $h(\varepsilon) = \varepsilon$.

Question 3

Soit $w \in \Sigma^*$. On cherche à montrer que $h(w)$ ne peut pas contenir de facteur qui soit un cube de longueur impaire. On raisonne par l'absurde : on suppose que $h(w)$ admet un facteur c^3 de longueur impaire.

1. Montrer que c^3 ne peut pas être un préfixe de $h(w)$.
2. Conclure.

Solution.

1. On considère les deux premières occurrences successives de c dans $h(w)$. Remarquons que la première commence en 0, et la seconde en $|c|$ (qui est impair). Quitte à étudier le cas symétrique, on suppose

$c[0] = a$. Alors, par construction de h, si on connaît une lettre en position paire de $h(w)$, on connaît celle à la position suivante ; donc $c[1] = b$. Donc il y a un b en position $|c| + 1$ (correspondant à la deuxième occurrence de c) dans $h(w)$; donc il y a un a après en position $|c| + 2$. Donc, en revenant à la première occurence, il y a un a en position 2. Donc il y a un b en position 3 ; et ainsi de suite. On en conclut que c est de la forme $(ab)^k a$. Or, cela signifie que l'image de $w[k]$ serait aa, ce qui est absurde !

2. On ne suppose plus que c^3 est un préfixe de $h(w)$. Quitte à retirer les premières lettres de w, on peut supposer que l'apparition de c^3 commence dans l'image de la première lettre de w. Par la question précédente, c^3 n'en est pas un préfixe ; donc $h(w)$ s'écrit sous le format $\ell c^3 \ldots$ avec ℓ une lettre. Remarquons alors que w peut s'écrire $p_w m$ où $h(p_w) = \ell c$ pour des raisons de parité. Donc c^2 est un préfixe de $h(m)$; notre raisonnement précédent invalide cette existence.

Question 4

Démontrer que pour tout mot $w \in \Sigma^*$, si $h(w)$ contient un facteur de longueur paire qui est un cube, alors w contient un facteur qui est un cube.

Solution.

De même, soit c^3 ce facteur cubique de $h(w)$. Quitte à oublier les premières lettres de w, on peut se ramener au cas où c^3 couvre l'image de la première lettre de $h(w)$.
Si c'est un préfixe de $h(w)$, alors pour des raisons de parité de la longueur, $h(w_{[\![0,|c|/2-1]\!]}) = c$; or, h est injective. Donc, en notant p ce préfixe de h, w commence par p^3.
Si $h(w) = \ell c^3 \ldots$, traitons le cas $\ell = a$. Alors $c[0] = b$; or, dans la deuxième occurrence de c, ce b se situe sur une position impaire, et juste avant, il y a $c[-1]$. Donc $c[-1] = a$. Donc c peut s'écrire comme $c = c'a$; donc $h(w)$ commence par $ac'ac'ac'a \ldots$, donc par le cube $(ac')^3$: on revient au cas précédent, et w contient un cube.

Question 5

Montrer que $(h^n(a))_{n\in\mathbb{N}}$ définit une famille de mots sans cube arbitrairement longs sur l'alphabet $\Sigma = \{a, b\}$.

Solution.

En posant $w_0 = a$ et $w_{n+1} = h(w_n)$, on remarque que par la question 3, pour tout $n \in \mathbb{N}$, w_n ne contient aucun cube impair ; puis, si w_n contenait un cube pair, alors w_{n-1} aussi, alors w_{n-2} aussi, ... Donc w_0 aussi, ce qui est absurde.
Donc aucun w_n ne contient de cube. Remarquons que la taille des w_n explose ($|w_n| = 2^n$).

Question 6

Un *pré-cube* est un mot de la forme $v \cdot v \cdot v[0]$ où $v \in \Sigma^*$ est non vide. Un mot est *sans pré-cube* si aucun de ses facteurs n'est un pré-cube. Montrer que $(h^n(a))_{n\in\mathbb{N}}$ définit une famille de mots sans pré-cube arbitrairement longs sur l'alphabet $\Sigma = \{a, b\}$.
On pourra dinstinguer les cas où v est de longueur impaire ou paire.

Solution.

Dans le cas où v est de longueur impaire : le raisonnement de la question 3 prouve qu'en fait, $h(w)$ ne peut jamais contenir de carré dont la racine est impaire. Donc il n'y aura pas de pré-cube avec v de longueur impaire.
Dans le cas de la longueur paire : on peut adapter le raisonnement de la question 4 pour prouver que si $h(w)$ contient un pré-cube, alors w aussi.
Si le pré-cube est un préfixe, alors v a exactement une pré-image u, et w doit commencer par deux occurrences de u. Remarquons que par définition de h, $u[0] = v[0]$. Pour la dernière lettre, toujours par parité, w doit

continuer sur $u[0]$. Donc w a pour préfixe un pré-cube.
Si $h(w) = \ell vvv[0]\dots$, alors si $\ell = a$, par parité la première lettre de v est un b, et donc la dernière est un a. Donc en notant $v = v'a$, $h(w) = av'av'av[0]\dots$, donc $h(w)$ commence par un pré-cube.
Le même raisonnement permet de conclure : la suite des w_n ne contient aucun pré-cube.

Question 7

Montrer qu'il existe des mots sans carré arbitrairement longs sur un alphabet de taille 4.

Solution.

On construit $x_n \in \Sigma^*$ de la façon suivante :

$$x_n[i] = \begin{cases} A \text{ si } w_n[i] = w_n[i+1] = a \\ B \text{ si } w_n[i] = w_n[i+1] = b \\ C \text{ si } w_n[i] = a \text{ et } w_n[i+1] = b \\ D \text{ si } w_n[i] = b \text{ et } w_n[i+1] = a \end{cases}$$

Si x_n contient un carré, alors w_n contient un pré-cube, ce qui n'est pas le cas. Donc $(x_n)_{n\in\mathbb{N}}$ est une famille infinie sans carré sur l'alphabet $\{A, B, C, D\}$.

Question 8

Montrer qu'il existe des mots sans carré arbitrairement longs sur un alphabet de taille 3.

Solution.

On construit $y_n \in \Sigma^*$ de la façon suivante :

$$x_n[i] = \begin{cases} E \text{ si } w_n[i] = w_n[i+1] \\ C \text{ si } w_n[i] = a \text{ et } w_n[i+1] = b \\ D \text{ si } w_n[i] = b \text{ et } w_n[i+1] = a \end{cases}$$

(autrement écrit, E confond A et B). Remarquons qu'on peut définir y_n comme l'image de x_n par un morphisme. Supposons que y_n contienne un carré vv.
Si $|v| = 1$, alors $v = EE$ (les autres cas sont couverts par la question précédente), mais cela impliquerait que w_n contienne trois lettres identiques d'affilée.
Si $|v| \geqslant 2$, alors v ne peut pas contenir EE (paragraphe précédent). On montre alors qu'en lisant de gauche à droite v, on peut deviner si le E était avant un A ou un B : en effet, si E est précédé d'un C, alors la première lettre couverte par E est un b, donc E était un B ; si E est précédé d'un D, alors E était un A. Donc la valeur recouverte par E ne dépend que des lettres à gauche et à droite de chaque E, et cela ne dépend que des lettres internes à v. Donc x_n contiendrait un carré, ce qui est absurde.

ZONE BONUS.

Remarque : **il n'est pas attendu que vous prépariez les questions suivantes**. Elles sont la suite présente dans le sujet original, et sont présentes au cas où vous ayez été extrêmement rapide ou pour vous entraîner.

Question 9

Un *mot infini* sur l'alphabet Σ est une séquence infinie $w[0] \dots w[n] \dots$ d'éléments de Σ. On dit qu'il est sans carré si aucun de ses facteurs finis n'est un carré, c'est-à-dire que pour tous $i < j$ dans N, le mot $w[i] \dots w[j-1]$ n'est pas un carré.
Existe-t-il des mots infinis sans cube sur un alphabet de taille 2 ? des mots infinis sans carré sur un alphabet de taille 3 ?

Solution.

Oui, il suffit de considérer les limites des suites $(w_n)_{n \in \mathbb{N}}$ et $(x_n)_{n \in \mathbb{N}}$ dans la bonne topologie.

Question 10

On construit une suite de mots en posant $w_0 := a$ et $w_{i+1} = h(w_i)$ pour tout $i \in \mathbb{N}$, où h est défini juste avant la question 2. Écrire le pseudo-code d'un algorithme qui, étant donné $n \in \mathbb{N}$, calcule w_n. Quelle est sa complexité en temps et en espace ?

Solution.

Remarquons que la suite des $(w_n)_{n \in \mathbb{N}}$ est préfixe : $w_n \sqsubseteq_p w_{n+1}$. Donc il suffit de considérer w_n ; de regarder les 2^{n-1} dernières lettres de w_n et de stocker leurs images successives à la fin de lui-même. Complexité spatiale minimale (en $O(2^n)$) et temporelle de même.

Remarque : je pense que le créateur du sujet n'avait pas remarqué cette propriété de préfixation, et pensait qu'on devait nécessairement repartir de 0 à chaque fois.

Question 11

Proposer une autre définition, par récurrence, de la suite (w_i) et écrire le pseudo-code d'un algorithme plus efficace pour calculer w_i étant donné i.

Solution.

Remarquons que si on note anti $: \Sigma^* \to \Sigma^*$ le morphisme défini par anti$(a) = b$, b $= a$, alors $w_{n+1} = w_n \cdot \text{anti}(w_n)$. De nouveau, on opère en temps et espace minimaux.