Mots panpermutationnels

27 octobre 2018

Définition 1. *Soit $A = \{1, \dots, n\}$ et A^* l'ensemble des mots sur l'alphabet A. Un mot $m \in A^*$ est dit panpermutationnel si pour toute permutation $\sigma \in \mathfrak{S}_n$, le mot $\sigma(1)\sigma(2)\dots\sigma(n)$ est un sous-mot de m, et on note A^*_{pp} l'ensemble des tels mots. Lorsque m est un mot, on notera $m_{[i:j]} = m_i m_{i+1} \dots m_j$.*

On s'intéresse à estimer la longueur minimale L_n d'un mot panpermutationnel.

Définition 2. *Soit un mot m panpermutationnel, on dit que le i-ème caractère m_i de m (pour $i \in \{1, \dots, |m| - n + 1\}$) est un caractère perdu si on est dans l'une des deux situations suivantes :*
- *Le sous mot $m_{[i:i+n-1]}$ n'est pas le code d'une permutation (i.e. il y a un caractère répété)*
- *Le sous mot $m_{[i:i+n-1]}$ est déjà apparu avant, c'est-à-dire qu'il existe $j < i$ tel que $m_{[j:j+n-1]} = m_{[i:i+n-1]}$.*

Il est alors clair que la longueur d'un mot $m \in A^*_{pp}$ est égale à $n! + n - 1 + L(m)$, où $L(m)$ est le nombre de caractères perdus de m, puisqu'il y a une bijection naturelle entre les caractères m_i de m (pour $i \in \{1, \dots, |m| - n + 1\}$) qui ne sont pas perdus et les permutations de $\mathfrak{S}_n$ (donnée par le code $m_{[i:i+n-1]}$). Le problème revient donc à estimer le nombre minimal de caractères perdus dans un mot panpermutationnel.

1 Borne inférieure

1.1 Une première borne

Lemme 3. *On perd toujours au moins $(n-1)! - 1$ caractères.*

Preuve. Pour prouver ce résultat, considérons une permutation $m_{[i:i+n-1]}$ apparaissant dans le mot panpermutationnel m à un certain indice i correspondant à un caractère non perdu. Si le $i+1$-ième caractère n'est pas perdu (c'est-à-dire qu'on ne perd aucun caractère entre les deux permutations), alors $m_{i+n} = m_i$, puisque $m_{[i+1:i+n]}$ doit être une permutation. Le mot commençant à l'indice $i+1$ est alors $m_{[i+1:i+n-1]}m_i$, c'est-à-dire la permutation circulaire (vers la gauche) de $m_{[i:i+n-1]}$

Autrement dit, on a montré que si deux permutations successives apparaissant dans le mot ne sont pas dans la même orbite pour la permutation circulaire, on perd nécessairement au moins un caractère entre elles.

Puisqu'on doit explorer toutes les orbites pour la permutation circulaire, qui sont au nombre de $\frac{n!}{n} = (n-1)!$, on est obligé de faire au moins $(n-1)! - 1$ changements d'orbites, et donc de perdre $(n-1)! - 1$ caractères. □

Ainsi $L_n \geq n! + (n-1)! + n - 2$.

Si $\sigma \in \mathfrak{S}_n$ est une permutation, on note $\bar{\sigma}$ la classe des permutations circulaires de σ (il y a $(n-1)!$ telles classes) et $L_{\bar{\sigma}}(m)$ la *contribution de la classe $\bar{\sigma}$ à la perte de caractères dans m*, c'est-à-dire le nombre de caractères c perdus de m tels que le premier caractère non-perdu précédant c est le début d'un sous-mot de longueur n qui code une permutation de $\bar{\sigma}$ (i.e. « les caractères qu'on a dû perdre pour sortir de la classe $\bar{\sigma}$ »). En notant $L_\infty(m)$ le nombre de caractères perdus au début de m, on a $L(m) = \sum_{\bar{\sigma}} L_{\bar{\sigma}}(m) + L_\infty(m)$.

Par exemple, la preuve du lemme précédent aurait pu se reformuler en disant que pour toute classe de permutations sauf éventuellement celle de la dernière permutation qui apparaît dans le mot, $L_{\bar{\sigma}}(m) \geq 1$. En effet, comme on l'a montré, on est obligé de perdre un caractère pour «recoller» une permutation à une permutation qui n'est pas dans la même classe.

1.2 Une borne meilleure

Lemme 4. *On perd toujours au moins $(n-2)! - 1$ caractères en plus de ceux déjà comptés, c'est-à-dire que $L(m) \geq (n-1)! + (n-2)! - 2$.*

Preuve. Pour prouver ce résultat, soit une classe $\bar{\sigma}$ de permutations de $\mathfrak{S}_n$. Soit $m_{[i:i+n-1]}$ l'unique sous-mot m (pour un certain indice i correspondant à un caractère non perdu) vérifiant la propriété suivante : $m_{[i:i+n-1]}$ code pour une permutation de $\bar{\sigma}$ et aucun autre élément de $\bar{\sigma}$ n'apparaît plus tôt dans le mot. Il y a deux possibilités (non exclusives) :

— Soit le sous-mot de longueur $2n-1$ commençant à l'indice i est $m_{[i:i+n-1]}m_{[i:i+n-2]}$.
— Soit il existe une permutation circulaire de $m_{[i:i+n-1]}$ qui apparaît plus tard dans le mot sans être contenue dans le sous-mot de longueur $2n-1$ commençant à l'indice i.

Dans le deuxième cas, sauf si la deuxième permutation conclut le mot, il est clair que $L_{\bar{\sigma}}(m) \geq 2$, puisqu'on a dû « sortir » deux fois de $\bar{\sigma}$ et qu'on sait que changer de classe de permutations fait toujours perdre un caractère (vu la preuve du lemme précédent). Si la deuxième permutation conclut le mot, ce qui arrive au plus pour une classe, on a tout de même $L_{\bar{\sigma}}(m) \geq 1$.

Dans le premier cas, sauf si le $i+i+2n$-ième caractère conclut le mot (auquel cas $L_{\bar{\sigma}}(m) = 0$), on voit déjà que le $i+n$-ième caractère (le deuxième m_i) est perdu, puisque soit il n'est pas suivi d'une permutation, soit il est suivi de $m_{[i:i+n-1]}$ qu'on a déjà vue. On sait donc déjà que $L_{\bar{\sigma}}(m) \geq 1$. Supposons désormais que le $i+n+1$-ième caractère (le deuxième m_{i+1}) n'est pas perdu, c'est-à-dire que $L_{\bar{\sigma}}(m) < 2$. Alors $m_{[i+1:i+n-2]}m_{i+2n-1}m_{i+2n}$ est une permutation et donc $\{m_{i+2n-1}, m_{i+2n}\} = \{m_i, m_{i+n-1}\}$.

Si $m_{i+2n-1} = m_{i+n-1}$ et $m_{i+2n} = m_i$, alors la permutation obtenue est une permutation circulaire de $m_{[i:i+n-1]}$. On l'a donc déjà vue comme sous-mot de $m_{[i:i+n-1]}m_{[i:i+n-2]}$ ce qui contredit que le $i+n+1$-ième caractère n'est pas perdu.

Ainsi $m_{i+2n-1} = m_i$ et $m_{i+2n} = m_{i+n-1}$, autrement dit la permutation suivante est obtenue en permutant circulairement (vers la gauche) les $n-1$ premiers éléments. Soit $T(\bar{\sigma})$ la classe de la permutation ainsi obtenue. Si elle est du deuxième type, alors (sauf si on est à la fin du mot) $L_{T(\bar{\sigma})} \geq 2$, sinon soit elle vérifie $L_{T(\bar{\sigma})} \geq 2$, soit on peut définir de même $T^2(\bar{\sigma})$, et ainsi de suite.

Il y a deux issues possibles à cette construction, puisqu'il y a un nombre fini de classes :
— Soit la suite $(T^k(\bar{\sigma}))$ boucle.
— Soit la suite $(T^k(\bar{\sigma}))$ termine.

Tant que la suite se poursuit, elle décrit un mot de la forme suivante :

$$m_{[i:i+n-1]}m_{[i:i+n-2]}m_i m_{i+n-1}m_{i+1}m_{[i+2:i+n-2]}\cdots$$

En effet, si en passant à la permutation suivante on trouve une permutation dont la classe est du premier type, c'est que la permutation trouvée (qui ne commence pas à un caractère perdu, par hypothèse) est la première de sa classe à apparaître dans le mot, et on peut alors recommencer le procédé.

Pour cette raison, il est clair que la suite d'entiers i_k formée par les indices des caractères non perdus qui codent la première permutation de chaque classe $T^k(\bar{\sigma})$ est strictement croissante ($i_{k+1} = i_k + n + 1$) et majorée par la taille du mot, et donc la suite doit s'arrêter.

Il est impossible de tomber sur une boucle.

On appelle finale une classe de permutations $\bar{\sigma}$ telle que $T(\bar{\sigma})$ n'est pas définie : soit $\bar{\sigma}$ est à la fin du mot (ce qui inclut le cas problématique pour les classes de permutations du deuxième type), soit on est dans un des deux cas où on a prouvé $L_{\bar{\sigma}}(m) \geq 2$. La suite $(T^k(\sigma))$ se termine nécessairement (pour toute classe de départ) sur une classe finale.

Tant que la suite dure, on a des sous-mots de la forme $\sigma_k(j)\sigma_k(j+1)\ldots\sigma_k(j+n-1)\sigma_k(j)\sigma_k(j+1)\ldots\sigma_k(j+n-2)$ qui s'enchaînent en se superposant, avec $\sigma_{k+1} = \sigma_k \circ g$ où :

$$g = \begin{pmatrix} 1 & 2 & 3 & \ldots & n-1 & n \\ 2 & 3 & 4 & \ldots & 1 & n \end{pmatrix}$$

En particulier : $T(\bar{\sigma}) = \overline{\sigma \circ g}$ si on prend pour σ la première permutation de la classe $\bar{\sigma}$ qui apparaît dans le mot. Et alors $\sigma \circ g$ est effectivement la première permutation de la classe $T(\bar{\sigma})$ qui apparaît dans le mot (sauf éventuellement pour la classe finale obtenue à la fin), et ainsi de suite, autrement dit : $T^k(\bar{\sigma}) = \overline{\sigma \circ g^k}$.

Mais $g^{n-1} = Id$, et donc la suite ne peut pas durer plus de $n-1$ étapes sans quoi elle bouclerait. Par ailleurs il est clair que $T(\sigma) = T(\sigma') \Rightarrow \sigma = \sigma'$. Donc, pour chaque classe finale $\bar{\tau}$, il y a au plus plus $n-1$ « antécédents » (classes $\bar{\sigma}$ vérifiant $\exists i$, $\bar{\tau} = T^i(\bar{\sigma})$). Puisque toute classe est l'antécédent d'une classe finale, on en déduit : $(n-1)! \leq (n-1)F$ où F est le nombre de classes finales. Ainsi $F \geq \frac{(n-1)!}{n-1} = (n-2)!$. On sait que toute classe finale, sauf potentiellement une, fait perdre deux caractères, et que toute classe non-finale fait perdre un caractère. Ainsi :

$$\begin{aligned} L(m) &= \sum_{\bar{\sigma}} L_{\bar{\sigma}}(m) + L_\infty(m) \\ &\geq \sum_{\bar{\sigma}} L_{\bar{\sigma}}(m) \\ &\geq \sum_{\bar{\sigma}\text{non finale}} 1 + \sum_{\bar{\sigma}\text{ finale qui ne termine pas le mot}} 2 \\ &\geq ((n-1)! - F) + (F-1)2 = (n-1)! + F - 2 \geq (n-1)! + (n-2)! - 2 \end{aligned}$$

Ainsi $L(m) \geq (n-1)! + (n-2)! - 2$

□

Cela donne la borne $L_n \geq n! + (n-1)! + (n-2)! + n - 3$.

1.3 Continuer la même démarche ?

2 ?

On appelle mot circulaire de générateur $m_{[i:i+n-1]}$ le mot $m_{[i:i+n-1]}m_{[i:i+n-2]}$.

On essaye de recoller des mots de la forme suivante (on met en rouge les caractères perdus et on alterne crochets et parenthèses en bleu pour indiquer les positions des mots circulaires successifs) :

$$[m_{[i:i+n-1]}m_i(m_{[i+1:i+n-2]}]m_im_{i+n-1}m_{i+1}[m_{[i+2:i+n-2]}m_i)m_{i+1}m_{i+n-1}m_{i+2}$$
$$(m_{[i+3:i+n-2]}]m_im_{i+1}]m_{i+2}m_{i+n-1}m_{i+3}[m_{[i+4:i+n-2]}m_{[i:i+2]})m_{i+3}m_{i+n-1}\dots$$
$$\dots m_{i+k}[m_{[i+k+1:i+n-2]}m_{[i:i+k-1]})m_{i+k}m_{i+n-1}m_{i+k+1}\dots$$
$$\dots m_{i+n-3}[m_{i+n-2}m_{[i:i+n-4]})m_{i+n-3}m_{i+n-1}m_{i+n-2}m_im_{i+1}m_{[i+2:i+n-3]}]$$

On nommera un tel mot « mot 2-circulaire de générateur $m_{[i:i+n-1]}$ ». Il est de longueur $(n+1)\times(n-1)+n-2 = n^2+n-3$.

On sait déjà qu'on est obligé de perdre deux caractères à la fin, à savoir le dernier m_{i+n-2} et le m_i qui le suit. Si on suppose que le m_{i+1} suivant n'est pas perdu (c'est-à-dire qu'on perd exactement deux caractères pour «recoller» les deux mots 2-circulaires), alors $m_{i+1}m_{[i+2:i+n-3]}m_?m_{?+1}m_{?+2}$ [1] est une permutation, donc $\{m_?, m_{?+1}, m_{?+2}\} = \{m_i, m_{i+n-2}, m_{i+n-1}\}$. Il faut exclure le cas où $m_{i+1}m_{[i+2:i+n-3]}m_?m_{?+1}m_{?+2}$ est une permutation circulaire d'une permutation obtenue en permutant circulairement les $n-1$ premiers éléments de $m_{[i:i+n-1]}$ (auquel cas on l'a déjà vue dans le mot 2-circulaire qu'on a considéré). Cette condition revient à exclure les valeurs suivantes de $(m_?, m_{?+1}, m_{?+2})$: $(m_{i+n-1}, m_{i+n-2}, m_i)$, $(m_{i+n-2}, m_{i+n-1}, m_i)$, $(m_{i+n-2}, m_i, m_{i+n-1})$.
Restent les valeurs suivantes de $(m_?, m_{?+1}, m_{?+2})$: $(m_{i+n-1}, m_i, m_{i+n-2})$, $(m_i, m_{i+n-1}, m_{i+n-2})$, $(m_i, m_{i+n-2}, m_{i+n-1})$.
On définit G le sous-monoïde (et donc sous-groupe) de $\mathfrak{S}_n$ engendré par les trois permutations suivantes :

$$g_1 = \begin{pmatrix} 1 & 2 & 3 & \dots & n-3 & n-2 & n-1 & n \\ 2 & 3 & 4 & \dots & n-2 & n & 1 & n-1 \end{pmatrix}$$

$$g_2 = \begin{pmatrix} 1 & 2 & 3 & \dots & n-3 & n-2 & n-1 & n \\ 2 & 3 & 4 & \dots & n-2 & 1 & n & n-1 \end{pmatrix}$$

$$g_3 = \begin{pmatrix} 1 & 2 & 3 & \dots & n-3 & n-2 & n-1 & n \\ 2 & 3 & 4 & \dots & n-2 & 1 & n-1 & n \end{pmatrix}$$

On montre que $G = \mathfrak{S}_n$.

1. J'ai perdu trace des indices précis, mais peu importe.