TD d'algorithmique avancée
TD 5 : plus longue sous-séquence commune

Jean-Michel Dischler et Frédéric Vivien

Problématique

Une séquence est une suite finie de symboles pris dans un ensemble fini. Si $u = \langle a_1, a_2, ..., a_n \rangle$ est une séquence, où a_1, a_2, ..., a_n sont des lettres, l'entier n est la *longueur* de u. Une séquence $v = \langle b_1, b_2, ..., b_m \rangle$ est une *sous-séquence* de $u = \langle a_1, a_2, ..., a_n \rangle$ s'il existe des entiers $i_1, i_2, ..., i_m$ $(1 \leq i_1 < i_2 < ... < i_m \leq n)$ tels que $b_k = a_{i_k}$ pour $k \in [1, m]$. Par exemple, $v = \langle B, C, D, B \rangle$ est une sous-séquence de $u = \langle A, B, C, B, D, A, B \rangle$ correspondant à la suite d'indice $\langle 2, 3, 5, 7 \rangle$.

Une séquence w est une *sous-séquence commune* aux séquences u et v si w est une sous-séquence de u et de v. Une sous-séquence commune est *maximale* ou est une *plus longue sous-séquence* si elle est de longueur maximale. Par exemple : les séquences $\langle B, C, B, A \rangle$ et $\langle B, D, A, B \rangle$ sont des plus longues sous-séquences communes de $\langle A, B, C, B, D, A, B \rangle$ et de $\langle B, D, C, A, B, A \rangle$.

Résolution par programmation dynamique

1. On cherche à déterminer la longueur d'une sous-séquence commune maximale à $u = \langle a_1, a_2, ..., a_n \rangle$ et $v = \langle b_1, b_2, ..., b_m \rangle$. On note $L(i, j)$ la longueur d'une sous-séquence commune maximale à $\langle a_1, a_2, ..., a_i \rangle$ et $\langle b_1, b_2, ..., b_j \rangle$ $(0 \leq j \leq m,\ 0 \leq i \leq n)$. Donnez une récurrence définissant $L(i, j)$. *Indication* : on pourra distinguer les cas $a_i = b_j$ et $a_i \neq b_j$.
2. Écrivez alors un algorithme récursif calculant la longueur de la plus longue sous-séquence commune de deux séquences.
3. Montrez que cet algorithme est de complexité au moins exponentielle dans le cas où les deux séquences n'ont pas d'éléments en commun.
4. Écrivez alors un algorithme suivant le paradigme de la programmation dynamique et calculant la longueur de la plus longue sous-séquence commune de deux séquences.
5. Quelle est la complexité de cet algorithme ?
6. Modifiez l'algorithme précédent pour que l'on puisse en plus construire *une* plus longue sous-séquence commune et affichez une telle sous-séquence.