

TD n 4

Programmation dynamique et mots tives

1 Plus longue sous-séquence commune

On considère deux séquences (textes formés de lettres) S et T . Leurs longueurs sont notés respectivement m et n . Une **sous-séquence** d'un texte est obtenue en enlevant certaines lettres du texte. Par exemple `al gori thmi que a` comme sous-séquence `gothi que`. Une sous-séquence commune à S et T est une séquence qui est sous-séquence des deux à la fois. On s'intéresse à trouver la plus longue sous-séquence commune (PLSSC), c'est-à-dire celle avec le plus de lettres.

Pour cela on construit un tableau $LPLSSC[i][j]$ (avec $1 \leq i \leq m$ et $1 \leq j \leq n$) tel que $LPLSSC[i][j]$ est la **longueur** de la plus longue sous-séquence

- commune à S et T
- terminant avant la i ème lettre de S (inclusivement)
- terminant avant la j ème lettre de T (inclusivement)

Par exemple si $S = \text{al gori thmi que}$ et $T = \text{programmation}$ on a $LPLSSC[4][3] = 1$ (o est PLSSC de `al go` et `prog`) et $LPLSSC[4][3] = 1$ (o est PLSSC de `al go` et `prog`) et $LPLSSC[13][13] = 4$ avec `orti`

Exercice 1 Notons $Q = Q[1] \dots Q[k]$ une plus longue sous-séquence commune à S et T . Montrer que

- Si $T[m] = S[n]$ alors $Q[k] = S[m] = T[n]$ et $Q[1..k-1]$ est la plus longue sous-séquence commune à $S[1..m-1]$ et à $T[1..n-1]$;
- Si $S[m] \neq T[n]$ et $Q[k] \neq S[m]$ alors Q est une plus longue sous-séquence commune à $S[1..m-1]$ et T ;
- Si $S[m] \neq T[n]$ et $Q[k] \neq T[n]$ alors Q est une plus longue sous-séquence commune à S et $T[1..n-1]$

Exercice 2 En déduire la formule de récurrence générale qui permet de calculer $LPLSSC[i][j]$ en fonction de $LPLSSC[i-1][j-1]$, $LPLSSC[i][j-1]$ et $LPLSSC[i-1][j]$. Ne pas oublier de donner son initialisation).

Exercice 3 Écrire le programme de programmation dynamique qui, étant donné deux séquences, calcule la longueur de d'une plus longue sous-séquence commune.

Exercice 4 Comment retrouver *une* longue(s) sous-séquence commune en mémorisant des informations complémentaires ?

2 Plus longue sous-séquence croissante

Les objets sur lesquels on travaille maintenant sont des séquences de nombres entiers. Il y a au moins deux nombres par séquence. Par exemple, une telle séquence (de longueur 9) est : $T = (11, 5, 2, 8, 7, 3, 1, 6, 4)$. On note de manière générale $S = (s_1, \dots, s_i, \dots, s_n)$ une séquence de longueur $n > 1$.

On appelle **sous-séquence croissante** (SSC) de S une sous-séquence dont :

- les éléments sont pris de gauche à droite dans S ;
- les éléments croissent strictement de gauche à droite.

On appelle **intervalle croissant** (IC) une sous-séquence croissante sans trou (contiguë)

Par exemple, (2,3,4) et (1,6) sont des SSC de $T = (11, 5, 2, 8, 7, 3, 1, 6, 4)$ et (1,6) est de plus un IC.

Le but de ces exercices est de trouver la longueur des plus longs IC et SSC d'une séquence quelconque S , notés respectivement $LIC(S)$ et $LSSC(S)$.

Par exemple, les plus longs IC de T sont (2, 8) et (1, 6), donc $LIC(T) = 2$. Les plus longues SSC de T sont (2, 3, 4) et (2, 3, 6), donc $LSSC(T) = 3$.

L'opération élémentaire de mesure de la complexité est la comparaison.

Exercice 5 Montrer que, pour toute séquence S , on a : $LSSC(S) \geq LIC(S)$.

Exercice 6 Donner le principe d'un algorithme en $O(n)$ pour calculer $LIC(S)$.

Exercice 7 Pourquoi un algorithme comme le précédent ne permet-il pas de calculer $LSSC(S)$ avec un algorithme en $O(n)$?

Exercice 8 Pour calculer $LSSC(S)$, on va remplir un tableau à deux lignes et n colonnes. La première ligne contient s_i (l'élément de rang i de S), la seconde l_i , qui est par définition la longueur de la plus longue sous-séquence de S dont le dernier élément est s_i . Dans notre exemple :

i	1	2	3	4	5	6	7	8	9
s_i	11	5	2	8	7	3	1	6	4
l_i	1	1	1	2	2	2	1	3	3

Montrer et expliquer la relation de récurrence suivante. Quelle valeur faut-il donner au terme Max quand on n'a aucun j tel que $0 < j < i$ pour lequel $s_j < s_i$?

$$l_1 = 1$$

$$l_i = 1 + \max_{\substack{0 < j < i \\ s_j < s_i}} l_j$$

Exercice 9 Ecrire le programme qui calcule $LSSC(S)$ pour toute séquence S de longueur n . Quel est son ordre de grandeur de complexité ?