Fiche TD 4 - Algorithmes gloutons

Exercice 1 - Couverture par intervalles

Décrire un algorithme efficace qui, étant donné un ensemble $\{x_1, \ldots, x_n\}$ de n points sur une droite, détermine le plus petit ensemble d'intervalles fermés de longueur 1 qui contient tous les points donnés. Prouver la correction de votre algorithme et donner sa complexité.

Exercice 2 - Codage de Huffman

Soit Σ un alphabet fini de cardinal au moins deux. Un *codage binaire* est une application injective de Σ dans l'ensemble des suites finies de 0 et de 1 (les images des lettres de Σ sont appelées *mots de code*). Il s'étend de manière naturelle par concaténation en une application définie sur l'ensemble Σ^* des mots sur Σ. Un codage est dit *de longueur fixe* si toutes les lettres dans Σ sont codées par des mots binaires de même longueur. Un codage est dit *préfixe* si aucun mot de code n'est préfixe d'un autre mot de code.

1 - Le décodage d'un codage de longueur fixe est unique. Montrer qu'il en est de même pour un codage préfixe.

2 - Représenter un codage préfixe par un arbre binaire dont les feuilles sont les lettres de l'alphabet.

On considère un texte dans lequel chaque lettre c apparaît avec une fréquence $f(c)$ non nulle. A chaque codage préfixe de ce texte, représenté par un arbre T, est associé un coût défini par:

$$B(T) = \sum_{c \in \Sigma} f(c) l_T(c)$$

où $l_T(c)$ est la taille du mot binaire codant c. Si $f(c)$ est exactement le nombre d'occurences de c dans le texte, alors $B(T)$ est le nombre de bits du codage du texte.

Un codage préfixe représenté par un arbre T est *optimal* si, pour ce texte, il minimise la fonction B. Montrer qu'à un codage préfixe optimal correspond un arbre binaire où tout nœud interne a deux fils. Montrer qu'un tel arbre a $|\Sigma|$ feuilles et $|\Sigma| - 1$ nœuds internes.

3 - Montrer qu'il existe un codage préfixe optimal pour lequel les deux lettres de plus faibles fréquences sont soeurs dans l'arbre (autrement dit leurs codes sont de même longueur et ne diffèrent que par le dernier bit) - *Propri'té de choix glouton.*

Etant donnés x et y les deux lettres de plus faibles fréquences dans Σ, on considère l'alphabet $\Sigma' = \Sigma - \{x, y\} + \{z\}$, où z est une nouvelle lettre à laquelle on donne la fréquence $f(z) = f(x) + f(y)$. Soit T' l'arbre d'un codage optimal pour Σ', montrer que l'arbre T obtenu à partir de T' en remplaçant la feuille associée à z par un nœud interne ayant x et y comme feuilles représente un codage optimal pour Σ - *Propriété de sous-structure optimale.*

4 - Donner un algorithme pour trouver un codage optimal et donner sa complexité. A titre d'exemple, trouver un codage optimal pour $\Sigma = \{a, b, c, d, e, g\}$ et $f(a) = 45$, $f(b) = 13$, $f(c) = 12$, $f(d) = 16$, $f(e) = 9$ et $f(g) = 5$.