Correction partielle du TD8

Exercice 3 - Complexité de SUBSET-SUM

Montrer que le problème suivant SUBSET-SUM est $\mathcal{NP}$-complet.

SUBSET-SUM

Entrée : un ensemble fini S d'entiers positifs et un entier objectif t.
Question : existe-t-il un sous-ensemble $S' \subseteq S$ tel que $\sum_{x \in S'} x = t$?

Indication : vous pouvez par exemple effectuer une réduction à partir de 3-SAT. A partir d'un ensemble de clauses $C_0, ..., C_{m-1}$ sur les variables $x_0, ..., x_{n-1}$, considérer S l'ensemble des entiers $v_i = 10^{m+i} + \sum_{j=0}^{m-1} b_{ij} 10^j$ et $v'_i = 10^{m+i} + \sum_{j=0}^{m-1} b'_{ij} 10^j$, $0 \leq i \leq n-1$, où b_{ij} (resp. b'_{ij}) vaut 1 si le littéral x_i (resp. $\overline{x_i}$) apparaît dans C_j et 0 sinon, et des entiers $s_j = 10^j$ et $s'_j = 2.10^j$, $0 \leq j \leq m-1$. Trouver alors un entier objectif t tel qu'il existe un sous-ensemble $S' \subseteq S$ de somme t si et seulement si l'ensemble initial de clauses est satisfiable. Conclure. Quels autres entiers auraient aussi marché ?

Correction

SUBSET-SUM est évidemment NP.
Pour montrer qu'il est NP-complet nous effectuons une réduction à partir de 3-SAT. Chaque clause du problème 3-SAT possède exactement 3 variables, et il y a au total n variables et m clauses. Nous transformons une instance de 3-SAT pour trouver une solution de SUBSET-SUM. Les entiers du problème SUBSET-SUM sont représentés en notation décimale, avec une colonne pour chacune des n variables et une colonne pour chacune des m clauses de la formule de 3-SAT.
En formalisant avec les notations de l'indication, on crée les entiers suivants:

- un entier v_i par variable x_i, qui possède un 1 dans la colonne correspondant à cette variable, et un 1 dans la colonne de chaque clause dans laquelle x_i apparaît;
- un entier v'_i par variable x_i, qui possède un 1 dans la colonne correspondant à cette variable, et un 1 dans la colonne de chaque clause dans laquelle $\overline{x_i}$ apparaît;
- un entier s_j par clause, possédant un 1 dans la colonne correspondant à cette clause;
- un entier s'_j par clause, possédant un 2 dans la colonne correspondant à cette clause;
- l'entier objectif t, possédant un 1 dans chaque colonne des variables, et un 4 dans chaque colonne des clauses.

Pour illustrer cela (avec une ré-indexation des variables et des clauses par rapport à l'indication de l'énoncé, par souci de clarté):

Entier	n variables x_1 x_2 ... x_n	m clauses C_1 C_2 ... C_m	Commentaires
v_1	1 0 ... 0	...1...1.....	1: clauses dans lesquelles x_1 apparaît
v_2	0 1 ... 0	..1..........	1: clauses dans lesquelles x_2 apparaît
v'_i	...1...	....1.....1...	1: colonne x_i 1: clauses dans lesquelles $\overline{x_i}$ apparaît
s_1	0 0 ... 0	1 0 ... 0	Uniquement 1 dans colonne C_1
s_i	0 0 ... 0	....1....	Uniquement 1 dans colonne C_i
s'_1	0 0 ... 0	2 0 ... 0	Uniquement 2 dans colonne C_1
s'_i	0 0 ... 0	....2....	Uniquement 2 dans colonne C_i
t	1 1 ... 1	4 4 ... 4	Entier objectif

Nous voulons maintenant prouver qu'il existe un sous-ensemble de ces entiers que l'on peut sommer pour atteindre l'entier objectif si et seulement si la formule 3-SAT peut être satisfaite. Nous étudions tour à tour les deux sens de l'équivalence. Remarquons que l'on somme les entiers par colonne et qu'il n'y a pas de problème de retenue vu que l'on n'a jamais un total de plus de $3 * 1 + 1 + 2 = 6$ par colonne (trois 1 pour les trois littéraux de la clause, 1 et 2 pour les s_i et s'_i)

1. Si 3-SAT possède une solution, nous prenons dans notre sous-ensemble les entiers v_i (resp. v'_i) correspondant aux variables ayant pour valeur vrai (resp. faux). La somme de ces entiers possède un 1 dans chaque colonne des variables, et entre 1 et 3 dans chaque colonne correspondant à une clause. En ajoutant à ces entiers certains des s_j et s'_j, on peut se ramener à 4 dans chaque colonne correspondant à une clause (on part de 1,2 ou 3 et on rajoute 1 et/ou 2). On atteint donc l'objectif. Si la formule n'est pas satisfaite, la somme des v_i et v'_i donnera au moins un 0 dans une colonne de clause (pour une clause n'étant pas satisfaite) et on ne pourra pas atteindre la valeur 4 vu que l'on rajoute au plus $1 + 2 = 3$.

2. Dans l'autre sens, si on atteint l'objectif, on *doit* avoir choisi un entier avec un 1 dans chaque colonne de variable, donc on a défini une instance de 3-SAT (x_i à vrai si l'on a choisi v_i, et à faux sinon). Vu que l'on a au final un 4 dans chaque colonne de clause, et qu'au plus 3 peut venir des s_i et s'_i, il y a nécessairement au moins un 1 qui vient des v_i ou v'_i, signifiant que pour notre instance la clause est satisfaite. Ceci étant valable pour chaque colonne de clause, la formule 3-SAT est satisfaite.

L'équivalence est donc prouvée. De plus, la réduction précédente est polynomiale car on passe de n variables et m clauses à $2n+2m$ entiers possédant $n+m$ chiffres. SUBSET-SUM est donc NP-complet.