Complexité - TD 03

Benjamin Bordais

03 Décembre 2020

On rappelle la définition du problème suivant (dit de $\mathsf{SUBSET-SUM}$) :
— ENTRÉE : un ensemble d'entiers $T = \{n_1, \dots, n_k\}$, un entier cible $t \in \mathbb{N}$
— SORTIE : il existe un sous-ensemble $S \subseteq T$ tel que $\sum_{n \in S} n = t$
Ce problème est NP-complet.

D'autre part, on donne ici les notions d'*image de Parikh* et du lemme d'Euler qui seront vus dans le prochain cours. Considérons un graphe orienté $G = (V, E)$. L'image de Parikh d'un chemin $\rho \in E^*$ est le vecteur $v : E \to \mathbb{N}$ qui compte les occurrences de chaque arête dans le graphe défini inductivement par $v_\epsilon[e] = 0$ et pour $e' \neq e$, on a $v_{\rho \cdot e'}[e] = v_\rho[e]$ et $v_{\rho \cdot e}[e] = v_\rho[e] + 1$. On définit de plus le support $Supp(v)$ d'un vecteur $v : E \to \mathbb{N}$ par $Supp(v) = \{e \in E \mid v(e) > 0\}$. Enfin, pour $E' \subseteq E$, on définit $G_{E'} = (V_{E'}, E')$ le graphe induit par E' tel que $V_{E'} = \{v \in V \mid \exists v' \in V,\ (v, v') \in E' \vee (v', v) \in E\}$.

On a alors le lemme suivant :

Lemme 1 (Euler) *Un vecteur $v : E \to \mathbb{N}$ est l'image de Parikh d'un chemin de s à t (avec $s \neq t$) si et seulement si :*
— *$G_{Supp(v)}$ est connexe ;*
— *$\sum_{(s,u) \in E} v(s, u) = 1 + \sum_{(s,u) \in E} v(u, s)$*
— *$\sum_{(u,t) \in E} v(u, t) = 1 + \sum_{(t,u) \in E} v(t, u)$*
— *Pour tout $w \neq s, t$, on a $\sum_{(u,w) \in E} v(u, w) = \sum_{(w,u) \in E} v(w, u)$*

Problèmes NP-complet On examine deux problèmes (le premier sera vu en cours) où montrer l'appartenance à NP est plus dur que montrer la NP-dureté.

Question 1 *Montrer que décider, dans un graphe $G = (V, E)$, si un vecteur $v : E \to \mathbb{N}$ est l'image de Parikh d'un chemin entre $s \in V$ et $t \in V$ peut s'effectuer en temps polynomial (et de manière déterministe) en la taille du graphe et des entiers apparaissant dans v.*

Solution :

On vérifie en temps polynomial que les conditions du lemme d'Euler sont remplies. Le graphe $G_{Supp(v)}$ est calculable en temps polynomial et décider de sa connexité peut également se faire en temps polynomial (à l'aide de parcours en largeur/profondeur pour déterminer s'il existe un chemin entre deux sommets). Les trois autres points consistent ensuite à effectuer des opérations arithmétiques en nombre polynomial.

Graphe pondéré positivement

Question 2 *Montrer que le problème suivant est NP-complet.*
GraphWeightedPositively
— *ENTRÉE : un graphe orienté $G = (V, E)$, pondéré positivement par $p : E \to \mathbb{N}$, deux sommets $s \neq t \in V$, un entier $a \in \mathbb{N}$*

— *SORTIE : il existe un chemin entre s et t de poids cumulé égal à a*

On rappelle que le poids cumulé $p(\rho)$ d'un chemin $\rho \in E^*$ est égal à $\sum_{e\in\rho} p(e)$.

Solution :

Ce problème est NP-dur. On effectue une réduction depuis SUBSET − SUM. Soit une instance $T = \{n_1, \ldots, n_k\}, t \in \mathbb{N}$ de ce problème. On construit (en espace logarithmique) le graphe $G = (V, E)$ avec $V = \{(x_i)_{0\leq i\leq k}, (a_i, o_i)_{1\leq i\leq k}\}$ et $E = \{(x_i, a_i), (x_i, o_i), (a_i, x_{i+1}), (o_i, x_{i+1}) \mid 0 \leq i \leq k-1\}$. De plus, on pose $p : E \to \mathbb{N}$ avec $p(u, v) = n_i$ si $y = a_i$, $p(u, v) = 0$ sinon. On prend alors $s = x_0$, $t = x_k$ et $a = t$. Ainsi, il existe un chemin entre s et t de poids a si et seulement si il existe un sous-ensemble de T de somme cumulé égal à t.

Montrons à présent que ce problème est dans NP. On peut d'abord remarquer que le long d'un chemin entre s et t de poids égal a, les poids cumulés intermédiaires sont tous entre 0 et a (étant donné que l'on n'a pas de poids négatifs). Ainsi, s'il existe un chemin entre s et t de poids a, il en existe un de taille au plus $(a+1)\cdot|V|$. En effet, si un chemin ρ de s à t de poids a passe deux fois par le même sommet u (i.e. $\rho = \rho' \cdot u \cdot \rho'' \cdot u \cdot \rho'''$ avec $\rho', \rho'', \rho''' \in E^*$), avec le même poids cumulé intermédiaire en u, le chemin $\rho' \cdot u \cdot \rho'''$ est de s à t de même poids a. Il faut remarquer que l'on ne peut pas deviner un chemin de taille $(a+1)\cdot|V|$ car cela est exponentiel en le nombre de bits nécessaire pour écrire a. En revanche, on peut deviner un vecteur $v : E \to \{0, \ldots, a+1\}$ (de taille polynomial en l'entrée). En effet, chaque sommet peut n'être vu qu'au plus $a+1$ fois dans un chemin de s à t de poids cumulé égal à a. On peut ensuite vérifier en temps polynomial qu'il correspond bien à un chemin entre s et t (ce qui peut être fait, d'après la question 1) et que le poids cumulé $p(v) = \sum_{e\in E} v(e)\cdot p(e)$ de v est égal à a.

Graphe pondéré négativement On souhaite à présent montrer que le poblème suivant est NP-complet.

GraphWeighted

— ENTRÉE : un graphe orienté $G = (V, E)$, pondéré par $p : E \to \mathbb{Z}$, deux sommets $s \neq t \in V$, un entier $a \in \mathbb{Z}$

— SORTIE : il existe un chemin entre s et t de poids cumulé égal à a

On considère donc un graphe $G = (V, E)$ pondéré par $p : E \to \mathbb{Z}$. On commence par définir ou rappeler quelques notions et notations qui seront utiles dans la suite de l'énoncé et dans vos solutions : Pour une arête $e = (u, v) \in E$, on note ${}^\bullet e$ pour u et $e^\bullet$ pour v. Un *chemin de longueur ℓ dans G* est un mot $\rho = e_1 \cdots e_\ell \in E^+$ composé de $\ell > 0$ arêtes de E et tel que ${e_{i-1}}^\bullet = {}^\bullet e_i$ pour tout $i = 2, \ldots, \ell$. Si $\rho = e_1 \cdots e_\ell$ est un chemin, les notations ${}^\bullet\rho$ et $\rho^\bullet$ désignent ${}^\bullet e_1$ et ${e_\ell}^\bullet$ respectivement. Comme précédemment, le poids $p(\rho)$ d'un chemin est la somme $\sum_{i=1}^{\ell} p(e_i)$ des poids de ses arêtes.

Un chemin $\rho = e_1 \ldots e_\ell \in E^+$ est un *cycle* si ${e_l}^\bullet = {}^\bullet e_1$ et le cycle est *élémentaire* si les sommets ${}^\bullet e_1, \ldots, {}^\bullet e_\ell$ sont tous distincts.

Question 3 *On dit qu'un chemin ρ est* factorisé en cycles *si ρ est écrit sous la forme*

$$\rho = \rho_0 \cdot \sigma_1^{k_1} \cdot \rho_1 \cdot \sigma_2^{k_2} \cdots \rho_{r-1} \cdot \sigma_r^{k_r} \cdot \rho_r$$

telle que les facteurs $\sigma_1, \ldots, \sigma_r$ sont des cycles élémentaires, les entiers $k_1, \ldots, k_r$ sont non nuls et les facteurs $\rho_0, \ldots, \rho_r$ n'ont aucun facteur qui soit un cycle. (La notation w^k avec $k \in \mathbb{N}$ dénote la concaténation de k copies de w, avec $w^0 = \epsilon$ et $w^{k+1} = w^k \cdot w$).

Montrez que tout chemin admet une factorisation en cycles.

Solution :

Par induction sur $|\rho|$.

• Si $\rho = (u, u)$ alors on prend $r = 1$, $\sigma_1 = \rho$, $k_1 = 1$ et $\rho_0 = \rho_1 = \epsilon$ en notant que la définition de factorisation en cycle (de "FC") spécifie que les ρ_i sont des facteurs de ρ, pas forcément des chemins, et donc peuvent être vides.

• Si $\rho = (u, v)$ avec $u \neq v$ on prend $r = 0$ et $\rho_0 = \rho$.

• Si $\rho = \rho' \cdot (u, v)$ alors on prend une FC de ρ', sous la forme $\rho' = \rho_0 \cdot \prod_{i=1}^{r}(\sigma_i^{k_i}\rho_i)$, qui existe par hypothèse d'induction. On considère alors $\rho_r \cdot e$. Si ce suffixe de ρ ne contient pas de cycle, on obtient une FC de ρ en remplaçant ρ_r par $\rho_r \cdot e$ dans la FC de ρ'. Si $\rho_r \cdot e$ contient un cycle alors, puisque ρ_r n'en contient pas, c'est que $e^\bullet$ coïncide avec un sommet visité par ρ_r. On écrit $\rho_r = \rho'_r \cdot \rho''_r$ tel que $\rho'^\bullet_r = e^\bullet$. On en tire une FC de $\rho_r \cdot e$ via $\rho_r \cdot e = \rho'_r(\rho''_r \cdot e)^1\epsilon$. En remplaçant ρ_r par cette FC dans la FC de ρ' on obtient une FC de ρ.

Question 4 *Montrez que si G admet un chemin de poids a allant de s à t alors il existe en particulier un tel chemin avec une factorisation en cycles $\rho_0 \cdot \sigma_1^{k_1} \cdot \rho_1 \cdot \sigma_2^{k_2} \cdots \rho_{r-1} \cdot \sigma_r^{k_r} \cdot \rho_r$ telle que les σ_i's aient tous des poids $p(\sigma_i)$ différents et aucun de poids nul.*

Solution :

Par induction sur $|\rho|$. Le cas $|\rho| = 1$ est trivial (comme à la question 3). Si $|\rho| > 1$ on l'écrit $\rho = \rho' \cdot e$ avec $e \in E$: par hypothèse d'induction il existe un chemin ρ_{ind} de même poids que ρ' et admettant une factorisation en cycle de poids distincts (une FCPD) $\rho_{ind} = \rho_0 \cdot \prod_{i=1}^{r}(\sigma_i^{k_i}\rho_i)$. Notons que $\rho_{ind} \cdot e$ est un chemin de s à t de même poids que $\rho' \cdot e = \rho$ et qu'il nous suffit de montrer l'existence d'une FCPD pour $\rho_{ind} \cdot e$.

Si $\rho_r \cdot e$ est sans cycle on obtient une FCPD de $\rho_{ind} \cdot e$ en remplaçant ρ_r par $\rho_r \cdot e$ dans la FCPD de ρ_{ind}. Si $\rho_r \cdot e$ contient un cycle alors comme ρ_r est sans cycle on sait que, comme à la question précédente, $\rho_r = \rho'_r \cdot \sigma_{r+1}$ avec ρ'_r sans cycle. On a alors trois cas :

1. Si le poids de σ_i est nul alors $\rho_0 \cdot \prod_{i=1}^{r-1}(\sigma_i^{k_i}\rho_i) \cdot \sigma_r^{k_r} \cdot \rho'_r$ est une FCPD de $\rho_{ind} \cdot e$ (on a supposé $r > 0$).

2. Si pour tout $i = \{1, \ldots, r\}$ le poids de σ_i est différent du poids de σ_{r+1} alors $\rho_0 \cdot \prod_{i=1}^{r-1}(\sigma_i^{k_i}\rho_i) \cdot \sigma_r^{k_r} \cdot \rho'_r \cdot \sigma_{r+1}^1 \cdot \epsilon$ est une FCPD de $\rho_{ind} \cdot e$ (on a supposé $r > 0$).

3. Sinon il existe un unique (par hypothèse d'induction) $0 < j \leq r$ tel que σ_j et σ_{r+1} soient de même poids. Dans ce cas, et en supposant $j < r$ pour simplifier l'écriture,

$$\rho_0 \cdot \Big(\prod_{i=1}^{j-1} \sigma_i^{k_i}\rho_i\Big) \cdot \sigma_j^{1+k_j} \cdot \rho_j \cdot \Big(\prod_{i=j+1}^{r-1} \sigma_i^{k_i}\rho_i\Big) \cdot \sigma_r^{k_r} \cdot \rho'_r$$

est la FCPD d'un chemin de s à t de même poids que $\rho' \cdot e$.

Question 5 *Pour $G = (V, E)$ et $p : E \to \mathbb{Z}$, on notera k le nombre $|V|$ de sommets, m le nombre $|E|$ d'arêtes, et $P = \max_{e \in E} |p(e)|$ le plus grand poids (en valeur absolue). Ainsi la donnée du graphe utilise un espace mémoire en $O(k + m\lceil \log_2(P) \rceil)$.*

Donnez un polynôme à quatre variables $Q(x_1, x_2, x_3, x_4)$ tel que pour toute instance $\langle G, u, v, a \rangle$, si G a un chemin de poids a reliant u à v alors il existe un tel chemin de longueur bornée par $Q(k, m, P, a)$.

Solution :

Les résultats sur les FCPD restent valides quand les poids sont des relatifs. On sait donc que s'il existe un chemin de s à t de poids a il en existe un admettant une FCPD

$\rho_0 \cdot \prod_{i=1}^{r}(\sigma_i^{k_i} \cdot \rho_i)$. On considère un chemin ρ de poids a et une factorisation qui minimisent le nombre total cumulé de fois où un cycle est vu $\sum_{i=1}^{r} k_i$. On a

$$\left(\sum_{0\leq i\leq r} p(\rho_i)\right) + \left(\sum_{i=1}^{r} k_i \cdot p(\sigma_i)\right) = a$$

.

Notons qu'un circuit élémentaire σ a au plus k arêtes et donc $-kP \leq p(\sigma) \leq kP$. On notera P' pour kP et on sait donc que $r \leq 2P' + 1$ puisque les cycles sont de poids distincts.

Étape 1. Pour ρ et sa FCPD $\rho_0 \cdot \prod_{i=1}^{r}(\sigma_i^{k_i} \cdot \rho_i)$, soit $I(\rho)$ l'ensemble (éventuellement vide) des indices i tels que σ_i soit de poids (strictement) positif, et $J(\rho)$ celui des indices i tels que $p(\sigma_i) < 0$. Montrons que l'on peut supposer que :

$$(\forall i \in I(\rho), k_i \leq P') \vee (\forall i \in J(\rho), k_i \leq P')$$

En effet, s'il existe un cycle négatif σ_i de poids p^- avec $k_i > P'$ et un cycle σ_j de poids positif p^+ avec $k_j > P'$ si on remplace k_i par $k'_i = k_i - p^+ \geq 0$ et k_j par $k'_j = k_j + p^- \geq 0$ dans la FCPD, on ne change pas le poids total :

$$\begin{aligned}
\sum_{n=1}^{r} k'_n \cdot p(\sigma_n) &= \sum_{n=1, n\neq i,j}^{r} k'_n \cdot p(\sigma_n) + k'_i \cdot p(\sigma_i) + k'_j \cdot p(\sigma_j) \\
&= \sum_{n=1, n\neq i,j}^{r} k_n \cdot p(\sigma_n) + (k_i - p^+) \cdot p^- + (k_j + p^-) \cdot p^+ \\
&= \sum_{n=1}^{r} k_n \cdot p(\sigma_n)
\end{aligned}$$

Cependant, le nombre total $\sum k_i$ de fois où un cycle est vu est diminué alors qu'il était supposé minimal.

Étape 2. Supposons grâce à l'étape précédente que $\forall i \in J(\rho), k_i \leq P'$ (le cas $\forall i \in I(\rho), k_i \leq P'$ est similaire). On sait donc que :

$$\sum_{i\in I(\rho)} k_i \cdot p(\sigma_i) = a - \left(\sum_{0\leq i\leq r} p(\rho_i)\right) - \left(\sum_{i\in J(\rho)} k_i \cdot p(\sigma_i)\right) .$$

Donc, puisque $p(\rho_i) \geq -P'$ tout comme $p(\sigma_i)$, et comme $k_i \leq P'$ quand $i \in J(\rho)$:

$$\sum_{i\in I(\rho)} k_i \cdot p(\sigma_i) \leq a + \left(\sum_{0\leq i\leq r} P'\right) + \left(\sum_{i\in J(\rho)} P' \cdot P'\right) .$$

D'où

$$\sum_{i\in I(\rho)} k_i \cdot p(\sigma_i) \leq Q'(a, k, P) .$$

On en déduit que (car $p(\sigma_i) \neq 0$ pour tout $0 \leq i \leq r$) $k_i \leq a + (2P' + 1)P' + P'^3 = Q'(a, k, P)$ pour tout $i \in I(\rho)$. Cette borne s'applique aussi quand $i \notin I(\rho)$ puisque $k_i \leq P'$ quand $i \in J(\rho)$.

La longueur du chemin ρ vérifie donc :

$$|\rho| = \left(\sum_{0\leq i\leq r} |\rho_i|\right) + \left(\sum_{i=1}^{r} k_i \cdot |\sigma_i|\right) \leq \left(\sum_{0\leq i\leq r} k\right) + \left(\sum_{i=1}^{r} k_i \cdot k\right)$$
$$\leq (r+1)\cdot k + r\cdot k\cdot Q'(a,k,P) \leq (2P+2)\cdot k + (2P+1)\cdot k\cdot Q'(a,k,P)$$

Question 6 *Montrer que le problème* GraphWeighted *est* NP*-complet.*

Solution :

Le problème est NP–dur puisqu'il l'est déjà quand les poids sont tous positifs ou nuls. Pour montrer qu'il est dans NP, il suffit de donner un algorithme non déterministe en temps polynômial. Comme pour le cas positif, on devine un vecteur v de nombres d'occurrences des arcs $e \in V$ tel que $\sum_{e\in E} v[e] \leq Q(k,m,P,a)$, c.-à-d. qu'il suffit de deviner un vecteur d'une taille bornée par un polynôme fixé de n, la taille de l'instance.

Attention, si k et m sont bornés par n, les valeurs P et a peuvent quant à elles être exponentielles : c'est la taille de leur représentations qui est bornée par n. Donc les valeurs de v ne sont pas bornées polynomialement en n mais la taille d'une représentation de v est en $O(n^2)$ puisque $v[e] \leq Q(k,m,P,a)$ pour chaque $e \in E$.

On vérifie alors que v est bien l'image de Parikh d'un chemin de u à v grâce aux conditions d'Euler et que $\sum_{e\in E} v[e].p(e) = a$.

Un problème coNP-complet Soit une classe $\mathcal{C}$ de problèmes de décision. La classe co$\mathcal{C}$ correspond à l'ensemble des langages L tel que $\bar{L} \in \mathcal{C}$: co$\mathcal{C} = \{\bar{L} \mid L \in \mathcal{C}\}$.

Question 7 *Supposons que le langage L soit complet pour la classe $\mathcal{C}$. Exhiber un langage complet pour la classe* co$\mathcal{C}$*.*

Solution :

Le langage $\bar{L}$ est complet pout la classe co$\mathcal{C}$. En effet, $\bar{L} \in$ co$\mathcal{C}$ par définition. De plus, soit $L' \in$ co$\mathcal{C}$. On a $\bar{L'} \in \mathcal{C}$. Comme L est complet pour la classe $\mathcal{C}$, $\bar{L'}$ peut se réduire à L. C'est-à-dire, il existe une fonction f calculable en espace logarithmique telle que $x \in \bar{L'} \Leftrightarrow f(x) \in L$. On a alors $x \in L' \Leftrightarrow \neg(x \in \bar{L'}) \Leftrightarrow \neg(f(x) \in L) \Leftrightarrow f(x) \in \bar{L}$. Ainsi, f est également une réduction de L' vers $\bar{L}$. Ainsi, $\bar{L}$ est dur pour la classe de complexité $\mathcal{C}$.

Question 8 *Prouver que le problème de décision suivant est* coNP*-complet :*

Tautology *:*

— *ENTRÉE : un formule propositionelle ϕ sous forme normale disjonctive*

— *SORTIE : toute valuation ν satisfait ϕ*

Solution :

Montrons tout d'abord que ce problème est dans coNP. Le problème $\overline{\text{Tautology}}$ peut s'écrire ainsi :

— ENTRÉE : un formule propositionnelle ϕ sous forme normale disjonctive

— SORTIE : il existe une valuation ν qui ne satisfait pas ϕ

On a $\overline{\text{Tautology}} \in$ NP car il suffit de deviner une valuation et de vérifier (en temps polynomial) que celle-ci ne satisfait pas la formule. Montrons à présent que Tautology est coNP-dur. On sait, ar la question précédente, que le problème $\overline{\text{SAT}}$ est coNP-complet avec $\overline{\text{SAT}}$:

— ENTRÉE : un formule propositionnelle ϕ sous forme normale conjonctive
— SORTIE : toute valuation ν ne satisfait pas ϕ

Pour une formule $\phi = \wedge_{i=1}^{n}(\vee_{j=1}^{a_i} l_i)$, on pose $f(\phi) = \neg\phi = \vee_{i=1}^{n}(\wedge_{j=1}^{a_i} \neg l_i)$. On a $f(\phi)$ sous forme normale disjonctive, calculable en espace logarithmique et ϕ n'est pas satisfiable si et seulement si $f(\phi)$ est une tautologie. C'est-à-dire, $\phi \in \overline{\mathsf{SAT}} \leftrightarrow f(\phi) \in \mathsf{Tautology}$.

Question 9 *Un problème* coNP*-complet est-il (a priori) dans* NP *?*

Solution :

Si un problème coNP-complet était dans NP, on aurait que NP = coNP (ce qui est un problème ouvert ce jour). En effet, soit L un problème coNP-complet dans NP. Pour tout problème $L' \in \mathsf{coNP}$, L' se réduit à L en espace logarithmique, et comme NP est stable par réduction logarithmique, on en déduit que $L' \in \mathsf{NP}$. Ainsi, $\mathsf{coNP} \subseteq \mathsf{NP}$. Pour $A \in \mathsf{NP}$, on a alors $\bar{A} \in \mathsf{coNP} \subseteq \mathsf{NP}$, donc $\bar{\bar{A}} = A \in \mathsf{coNP}$. C'est-à-dire, $\mathsf{NP} \subseteq \mathsf{coNP}$. Ainsi, on a $\mathsf{coNP} = \mathsf{NP}$.