Calculabilité / Complexité (L3)
Devoir à la maison décembre 2016
Énoncés et solutions*

Exercice 1 : UNIQ_SAT

Le problème UNIQ_SAT est une variante de 3_SAT. On rappelle que 3_SAT considère des formules propositionnelles en forme normale conjonctive (CNF) telles que chaque clause contient exactement trois littéraux. Une instance a la forme $\psi = C_1 \wedge C_2 \wedge \cdots \wedge C_m$, où chaque clause C_i est de la forme $l_{i,1} \vee l_{i,2} \vee l_{i,3}$ et où chaque littéral $l_{i,j}$ est soit une variable propositionnelle $p \in \mathit{Prop} = \{p, q, r, \ldots\}$, soit sa négation $\neg p$. On notera $l_{i,j} = \epsilon_{i,j} p_{i,j}$ avec $\epsilon_{i,j} \in \{+, -\}$.

Alors que 3_SAT est le problème de savoir s'il existe une valuation $v : \mathit{Prop} \to \mathit{Bool}$ qui valide ψ, UNIQ_SAT demande s'il existe une valuation qui valide *un et un seul* littéral dans chaque clause.

NB. Attention, dans cet énoncé le nom UNIQ_SAT a été choisi pour rendre plus difficile une recherche de solutions sur internet. La variante étudiée s'appelle classiquement 1_in_3_SAT tandis que UNIQ_SAT désigne un autre problème qui n'est pas NP–complet.

Question 1. Donnez une instance ψ qui soit positive pour 3_SAT et négative pour UNIQ_SAT. Justifiez.

Solution. Le plus simple est $\psi = p \vee p \vee p$ (justification omise).

Question 2. Montrez que UNIQ_SAT est NP–complet. Justifiez la correction de votre réduction. Pour cette question, les seules réductions autorisées partent de problèmes vus en classe.

Solution. UNIQ_SAT est clairement dans NP (on devine une valuation v, puis pour chaque clause on vérifie l'existence *et l'unicité* d'un littéral validé par v). Pour montrer qu'il est NP–difficile, on donne une réduction de 3_SAT (déjà connu pour être NP–complet) dans UNIQ_SAT. Soit $\psi = \bigwedge_{i=1}^{m} l_{i,1} \vee l_{i,2} \vee l_{i,3}$ une instance de 3_SAT, on lui associe une instance ψ' obtenue à partir de ψ en remplaçant chaque clause C_i par les 3 clauses C_i', C_i'', et C_i''', suivantes:

$$(a_i \vee \neg l_{i,1} \vee b_i) \wedge (b_i \vee l_{i,2} \vee c_i) \wedge (c_i \vee \neg l_{i,3} \vee d_i) \,.$$

La réduction donnée plus haut est bien logspace (en lisant C_i on écrit directement $C_i' \wedge C_i'' \wedge C_i''$ et il suffit de maintenir un compteur, p.ex. i, pour engendrer les nouvelles variables propositionnelles a_i, b_i, c_i, d_i).

Pour montrer la correction, on suppose qu'il existe une valuation v qui valide ψ et on montre qu'on peut l'étendre sur $\{a_i, b_i, c_i, d_i\}_{i=1,\ldots,m}$ de façon à valider un unique littéral dans chaque clause de ψ'. Pour la clause C_i, si $v(l_{i,2}) = \mathit{vrai}$ alors on pose $v'(b_i) = v'(c_i) = \mathit{faux}$, $v'(a_i) = v(l_{i,1})$ et $v'(d_i) = v(l_{i,3})$. Si $v(l_{i,2}) = \mathit{faux}$ alors $v(l_{i,1} \vee l_{i,3}) = \mathit{vrai}$ et on peut supposer que p.ex. $v(l_{i,1}) = \mathit{vrai}$ (l'autre cas est symétrique). On peut alors poser $v'(b_1) = \mathit{vrai}$, $v'(c_i) = v'(a_i) = \mathit{faux}$ et $v'(d_i) = v(l_{i,3})$.

Réciproquement, si une valuation v valide exactement un littéral par clause de ψ' alors on montre qu'elle valide ψ. En effet, si $v(l_{i,1} \vee l_{i,3}) = \mathit{vrai}$ alors v valide C_i. Sinon $v(\neg l_{i,1}) = v(\neg l_{i,3}) = \mathit{vrai}$ et donc $v(b_i) = v(c_i) = \mathit{faux}$ puisque v ne valide qu'un littéral de C_i' et de C_i'''. Donc $v(l_{i,2}) = \mathit{vrai}$ puisque v valide C_i'' et donc v valide aussi C_i.

*Merci de signaler toute erreur ou typo à `phs@lsv.fr`.

Exercice 2 : SUBSETSUM

Le problème SUBSETSUM n'a pas été présenté en classe mais il est décrit dans le polycopié du cours. Il vous faut donc étudier les preuves des Propositions 7 et 8 (p. 20 *& seq.* du polycopié) où il est montré que SUBSETSUM est NP–complet quand les nombres $v_1, \ldots, v_n, w$ qui composent une instance sont donnés en binaire, et qu'il est dans P quand ces nombres sont donnés en base 1.

Question 1. On considère une version de SUBSETSUM où l'input consiste en un entier d (la dimension) suivi de vecteurs $v_1, \ldots, v_n, w$ de $\mathbb{N}^d$ et où on doit décider s'il existe un sous ensemble $I \subseteq \{1, 2, \ldots, n\}$ tel que $w = \sum_{i \in I} v_i$.

Pour ce problème, dénommé VECSUBSETSUM_unary, la dimension d ainsi que les vecteurs sont donnés en base 1, c.-à-d. que la taille de l'input est en $O\big(d + \sum_{j=1}^{d} w[i] + \sum_{i=1}^{n} \sum_{j=1}^{d} v_i[j]\big)$.

Est-ce que ce problème est dans P ou bien est-il NP–difficile ? Justifiez précisément votre réponse.

Solution. Le problème est NP–difficile (et donc NP–complet car il est évidemment dans NP). On peut le montrer par une réduction UNIQ_SAT $\leq$ VECSUBSETSUM_unary. Pour cette réduction on notera $\mathbf{0}_s$ le vecteur $\langle 0\,0 \cdots 0 \rangle$ composé de s zéros, $\mathbf{1}_t$ le vecteur de composé de t uns, et $v \cdot v'$ le vecteur obtenu en concaténant les vecteurs v et v'.
Soit une instance $\psi = C_1 \wedge \cdots \wedge C_m$ de UNIQ_SAT où les propositions utilisées sont dans $Prop = \{p_1, \ldots, p_k\}$. À ψ notre réduction associe une instance de dimension $d = 3k + m$ et composée de $4k$ vecteurs. Pour un littéral l on définit U_l comme étant le vecteur de dimension m tel que $U_l[j] = 1$ si l valide C_j, $= 0$ sinon. Définissons alors, pour une valuation v, $U_v = \sum_{i=1}^{k} U_{v(p_i)}$: ce vecteur compte, dans chaque clause C_j, le nombre de littéraux validés. Donc v valide un et un seul littéral par clause de ψ ssi $U_v = \langle 111 \cdots 1 \rangle = \mathbf{1}_m$.
Pour une proposition p_i, $i = 1, \ldots, k$, on considère les vecteurs $v_i^+, v_i^-, v_i^0, v_i^{0\prime}$ donnés par:

$$v_i^+ = \mathbf{0}_{3(i-1)} \cdot \langle 1\,0\,1 \rangle \cdot \mathbf{0}_{3(k-i)} \cdot U_{p_i} \qquad v_i^- = \mathbf{0}_{3(i-1)} \cdot \langle 0\,1\,1 \rangle \cdot \mathbf{0}_{3(k-i)} \cdot U_{\neg p_i}$$

$$v_i^0 = \mathbf{0}_{3(i-1)} \cdot \langle 1\,0\,0 \rangle \cdot \mathbf{0}_{3(k-i)} \cdot \mathbf{0}_m \qquad v_i^{0\prime} = \mathbf{0}_{3(i-1)} \cdot \langle 0\,1\,0 \rangle \cdot \mathbf{0}_{3(k-i)} \cdot \mathbf{0}_m$$

et on fixe l'objectif $w = \mathbf{1}_{3k+m}$. Si un sous-ensemble I de $\{v_i^+, v_i^-, v_i^0, v_i^{0\prime}\}_{i=1,\ldots,k}$ est de somme w alors I contient forcément un et un seul vecteur parmi chaque paire $\{v_i^+, v_i^-\}$, seule façon de garantir $w[3i] = 1$. Donc I correspond à une valuation v qui satisfait un et un seul littéral par clause puisque $\sum_{x \in I} x = \mathbf{1}_{3k+m}$ implique que $U_v = \mathbf{1}_m$. Donc v prouve que ψ est une instance positive de UNIQ_SAT. Réciproquement si ψ est une instance positive validée par une valuation v, la somme w des vecteurs correspondant à v est un vecteur de 1's sauf à certaines positions de la forme $3i - 1$ ou $3i - 2$. On complète alors avec des vecteurs v_i^0 ou $v_i^{0\prime}$ pour obtenir $\mathbf{1}_{3k+m}$.

Problème : chemins dans les graphes pondérés

On considère des graphes pondérés de la forme $G = (V, E, p)$ où les arêtes de $E \subseteq V \times V$ sont orientées et portent chacune un poids, un entier naturel donné par $p : E \to \mathbb{N}$.

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. 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 ${e_\ell}^\bullet = {}^\bullet e_1$ et le cycle est *élémentaire* si les sommets ${}^\bullet e_1, \ldots, {}^\bullet e_\ell$ sont tous distincts.

Le problème WEIGHTEDPATH a comme input un graphe pondéré G, deux sommets $u, v \in E$, un poids $a \in \mathbb{N}$. Il s'agit de décider s'il existe dans G un chemin allant de u à v et de poids total a.

Ce problème n'a pas été étudié en classe mais il est montré dans le polycopié du cours qu'il est dans NP (Proposition 10 p. 24, s'appuyant sur le lemme d'Euler).

Question 1. Lisez attentivement dans le polycopié la preuve de l'appartenance à NP. Redonnez, en la détaillant, une preuve que s'il existe un chemin de poids a allant de u à v alors il existe en particulier un tel chemin de longueur au plus $(a+1)|V|$.

Solution. Soit ρ un chemin de poids a allant de u à v. Si $|\rho| > (a+1)|V|$ alors on factorise ρ sous la forme $\rho = (\prod_{i=0}^{a} \rho_i) \cdot \rho'$ avec $|\rho_i| = |V|$ pour chaque i, c.-à-d. qu'on repère $a+1$ facteurs de longueur $|V|$ et qu'on garde le reste dans ρ'. Notons qu'un facteur $\rho_i = e_1.e_2 \cdots e_{|V|}$ de longueur $|V|$ visite $|V|+1$ sommets donc contient un cycle.
Puisque $a = p(\rho) = \sum_{i=0}^{a} p(\rho_i) + p(\rho')$, il existe nécessairement (au moins) un indice $k \in \{0, 1, \ldots, a\}$ tel quel $p(\rho_k) = 0$ et donc tel que toutes les arêtes de ρ_k soient de poids nul. En retirant de ρ_k un des cycles (forcément de poids nul) qu'il contient forcément, on obtient un chemin de u à v, plus court que ρ, et de poids inchangé.
Puisque on peut raccourcir ρ si sa longueur dépasse $(a+1)|V|$ on finit par construire un chemin de longueur au plus $(a+1)|V|$.

Question 2. On dit qu'un chemin ρ est *factorisé en cycles* si ρ est écrit sous la forme

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

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") dit 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 hyp. ind. 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 3. 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 $\rho_0 \sigma_1^{k_1} \rho_1 \sigma_2^{k_2} \cdots \rho_{r-1} \sigma_r^{k_r} \rho_r$ telle que les σ_i's aient tous des poids $p(\sigma_i)$ différents.

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 deux cas:
1. 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 $r > 0$).
2. Sinon il existe un $0 < j \leq r$ 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 4. On s'intéresse maintenant à des graphes où les poids sont des entiers relatifs. Pour $G = (V, E, p)$ avec $p : E \to \mathbb{Z}$, 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 et une factorisation qui minimisent $\sum_{i=1}^{r} k_i$.
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 $k_j > P'$ alors on peut remplacer k_i par $k_i - p^+$ et k_j par $k_j + p^-$ dans la FCPD sans changer le poids total tout en diminuant $\sum k_i$ qui 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:

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

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

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

D'où

$$\sum_{i \in I(\rho)} k_i.p(\sigma_i) \leq a + (2P' + 1)P' + P'^3 .$$

On en déduit que $k_i \leq a + (2P' + 1)P' + P'^3$ pour tout $i \in I(\rho)$. Cette borne s'applique aussi quand $i \notin I(\rho)$ puisque $k_i \leq P'$ quand $i \in J(\rho)$, et puisque on peut supposer $k_i = 1$ s'il y a un cycle σ_i de poids nul. La longueur du chemin est donc bornée par $r.k.Q'(a, k, P) + (r + 1)k$.

Question 5. Quelle est la complexité de WEIGHTEDPATH quand les poids sont des entiers relatifs?

Solution. Le problème est NP–complet. Il est évidemment NP–difficile puisqu'il l'est déjà quand les poids sont tous positifs. 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. On vérifie aussi que $\sum_{e \in E} v[e].p(e) = a$. Ces vérifications impliquent des calculs arithmétiques simples sur un nombre polynomial de valeurs qui s'écrivent toutes avec un nombre polynomial de chiffres, elles sont donc réalisables en temps polynomial.