Αλγόριθμοι και πολυπλοκότητα

4η Σειρά Γραπτών ασκήσεων

Χειμερινό Εξάμηνο 2013-2014 Μπογιόκας Δημήτριος - ΜΠΛΑ

Άσκηση 1

Άσκηση 2

Άσκηση 3

Άσκηση 4(α) Αρχικά ισχύει ότι $3Partition \in NP$. Πράγματι, σε μια δεδομένη διαμέριση χρειάζεται γραμμικός χρόνος για να σχηματισθούν τα επί μέρους αθροίσματα και να ελεγχθεί το κατά πόσον ισούνται μεταξύ τους. Στη συνέχεια θα δείξω ότι η 3-διαμέριση είναι NP-hard, δείχνοντας ότι

$$Partition \leq_m^p 3Partition$$

Έστω $\mathcal{A} = \{w_1, w_2, \ldots, w_n\}$, όπου $n \in \mathbb{N}$ $w_1, \ldots, w_n \in \mathbb{Z}_{>0}$ και $2 | \sum_{i=1}^n w_i$ μία είσοδος του προβλήματος $Partition$. Ορίζω τη συνάρτηση f ως εξής:

$$f(\mathcal{A}) := \left\{ w_1, w_2, \ldots, w_n, \frac{\sum_{i=1}^n w_i}{2} \right\}$$

για όλες τις πιθανές εισόδους $\mathcal{A}$. Τότε, το $f(\mathcal{A})$ είναι είσοδος του $3Partition$, χρειάζεται πολυωνυμικό χρόνο για να υπολογισθεί και επιπλέον ισχύει

$$\mathcal{A} \in Partition \Leftrightarrow f(\mathcal{A}) \in 3Partition$$

Πράγματι, έστω $\{w_1, \ldots, w_n\} \in Partition$. Δηλαδή, υπάρχουν ξένα υποσύνολα δεικτών $I, J \subseteq \{1, \ldots, n\}$ με $I \cup J = \{1, \ldots, n\}$ τέτοια ώστε:

$$\sum_{i \in I} w_i = \sum_{j \in J} w_j = b$$

όπου $b = \frac{1}{2} \sum_{i=1}^n w_i$. Άρα τα σύνολα $\{w_i : i \in I\}, \{w_j : j \in J\}, \{b\}$ αποτελούν προφανώς τριμέριση του $f(\{w_1, \ldots, w_n\})$ με τις επιθυμητές ιδιότητες. Άρα $f(\{w_1, \ldots, w_n\}) \in 3Partition$. Αντίστροφα έστω κάποιο $\mathcal{A} = \{w_1, \ldots, w_n\}$ τέτοιο ώστε $f(\mathcal{A}) \in 3Partition$. Τότε, αν b όπως πριν, κάθε ένα από τα 3 σύνολα της διαμέρισης έχει άθροισμα b και αφού το b είναι στοιχείο του αρχικού συνόλου, ένα από τα τρία σύνολα πρέπει να είναι το μονοσύνολο $\{b\}$. Άρα υπάρχουν υποσύνολα δεικτών I, J όπως πρίν, ώστε

$$\sum_{i \in I} w_i = \sum_{j \in J} w_j = b$$

άρα, προφανώς $\{w_1, \ldots, w_n\} \in Partition$.

Άσκηση 4(β) Αρχικά ισχύει ότι το πρόβλημα (θα το συμβολίζω με $STLeafs$) ανήκει στο NP. Πράγματι, ένα δεδομένο υποσύνολο ακμών μπορεί να ελεγχθεί σε πολυωνυμικό χρόνο ως προς το αν σχηματίζει spanning tree και κατά πόσον τα φύλλα αυτού είναι υποσύνολο δεδομένου συνόλου κορυφών L. Στη συνέχεια θα δείξω ότι το $STLeafs$ είναι NP-hard, δείχνοντας ότι

$$HamPath \leq_m^p STLeafs$$

όπου $HamPath$ είναι το πρόβλημα ύπαρξης μονοπατιού Hamilton. Έστω $G = (V, E)$ μια είσοδος του προβλήματος $HamPath$. Ορίζω τη συνάρτηση f ως εξής:

$$f(V, E) = \{G' = (V', E'), L = \{u_1, u_2\}\}$$

όπου

$$V' = V \cup \{v_1, v_2, u_1, u_2\}$$

και

$$E' = E \cup \{\{v_1, v\} : v \in E\} \cup \{\{v_2, v\} : v \in E\} \cup \{\{v_1, u_1\}, \{v_2, u_2\}\}$$

θεωρώντας ότι οι κορυφές u_1, u_2, v_1, v_2 που προσθέτω δεν ανήκουν στο αρχικό V. Ουσιαστικά η συνάρτηση παίρνει ως είσοδο ένα γράφημα, του προσθέτει αρχικά δύο κορυφές v_1, v_2 που τις ενώνει με όλες τις κορυφές του αρχικού γραφήματος και στη συνέχεια προσθέτει δύο κορυφές βαθμού 1, τις u_1, u_2 που τις ενώνει μόνο με την v_1 και τη v_2 αντιστοίχως. Τέλος, η είσοδος του $STLeafs$ αποτελείται από το G' και από το $L = \{u_1, u_2\} \subseteq V'$. Η f προφανώς είναι πολυωνυμικά υπολογίσιμη και επιπλέον ισχύει

$$G \in HamPath \Leftrightarrow f(G) \in STLeafs$$

Πράγματι, έστω G γράφημα, τέτοιο ώστε $G \in HamPath$. Δηλαδή, υπάρχουν κορυφές $s, t \in V$ και σύνολο ακμών $K = \{\{s = r_1, r_2\}, \{r_2, r_3\}, \ldots, \{r_{n-1}, r_n = t\}\}$ με $r_i \neq r_j$ όταν $i \neq j$ και επιπλέον $\cup_{i=1}^{n}\{r_i\} = V$. Σχηματίζω το σύνολο $K' \subseteq E'$ ως εξής:

$$K' = K \cup \{\{s, v_1\}, \{t, v_2\}, \{v_1, u_1\}, \{v_2, u_2\}\}$$

Το $G'[K']$ (το υπογράφημα του G' που επάγεται από το K' είναι προφανώς spanning tree του $G' = f(G)$ και κάθε κορυφή στο $G'[K']$, εκ κατασκευής, έχει βαθμό 2, εκτός από τις u_1, u_2 που είναι άρα τα μοναδικά φύλλα του δέντρου. Δηλαδή ισχύει $\{u_1, u_2\} \subseteq L$ (για την ακρίβεια τα σύνολα ταυτίζονται). Άρα $f(G) \in STLeafs$. Αντίστροφα, έστω G γράφημα, τέτοιο ώστε $G' = f(G) \in STLeafs$, δηλαδή υπάρχει spanning tree T του G', τέτοιο ώστε τα φύλλα του να είναι υποσύνολο του $\{u_1, u_2\}$. Αφού τα u_1, u_2 έχουν βαθμό 1 στο G', θα έχουν βαθμό ≤ 1 και σε κάθε υπογράφημα, ειδικότερα στο εν λόγω spanning tree. Δηλαδή τα φύλλα του T είναι ακριβώς οι κορυφές u_1, u_2. Είναι αρκετά εύκολο να δείξει κάποιος επαγωγικά ότι σε ένα δέντρο ισχύει ότι το πλήθος των φύλλων είναι τουλάχιστον τόσο όσο ο βαθμός της μέγιστης κορυφής. Αφού το T έχει δύο φύλλα, κάθε κορυφή του T έχει βαθμό το πολύ 2. Άρα, κάθε κορυφή του T, εκτός από τις u_1, u_2 έχει βαθμό ακριβώς 2. Δηλαδή, το T είναι Hamilton Path στο G'. Άρα, αν s, t είναι οι γείτονες των v_1 και v_2, αντιστοίχως, μέσα στο G, τότε το $T \setminus \{\{v_1, u_1\}, \{v_2, u_2\}, \{v_1, s\}, \{v_2, t\}\}$ είναι Hamilton Path στο G, δηλαδή $G \in HamPath$.

Άσκηση 4(γ) Αρχικά ισχύει ότι το πρόβλημα (έστω $FVSD$) ανήκει στο NP. Πράγματι, δεδομένου ενός υποσυνόλου S του συνόλου των κορυφών ενός γραφήματος $G = (V, E)$, σε πολυωνυμικό χρόνο μπορεί να διαπιστωθεί κατά πόσον το $G[V \setminus S]$ είναι ακυκλικό, μέσω του αλγορίθμου DFS. Στη συνέχεια θα δείξω ότι το $FVSD$ είναι NP-hard, δείχνοντας ότι

$$VC \leq_m^p FVSD$$

Έστω G ένα μη κατευθυνόμενο γράφημα και (G, k) μία είσοδος του προβλήματος VC. Ορίζω τη συνάρτηση f ως εξής:

$$f((V, E), k) = (G', k)$$

όπου $G' = (V', E')$ ένα κατευθυνόμενο γράφημα, τέτοιο ώστε:

$$V' = V \qquad\qquad E' = \{(u, v) : \{u, v\} \in E\}$$

Παρατηρώ ότι με τον παραπάνω συμβολισμό που χρησιμοποιώ, εννοείται ότι αν $\{u, v\} \in E$, τότε τόσο η (u, v), όσο και η (v, u) ανήκουν στο E'. Ουσιαστικά, το γράφημα G' παράγεται από το G αντικαθιστώντας κάθε ακμή $\{u, v\}$ του G με τις (κατευθυνόμενες) ακμές (u, v) και (v, u). Η f προφανώς είναι υπολογίσιμη σε γραμμικό χρόνο και επιπλέον ισχύει:

$$(G, k) \in VC \Leftrightarrow f(G, k) \in FVSD$$

Πράγματι, έστω γράφημα $G = (V, E)$ και $k \in \mathbb{Z}_{>0}$ τέτοια ώστε $(G, k) \in VC$, υπάρχει δηλαδή $K \subseteq V$, υπούνολο των κορυφών του G, τέτοιο ώστε $|K| \leq k$ και για κάθε ακμή

e του G να ισχύει $e \cap K \neq \emptyset$. Τότε, προφανώς το K έχει την επιθυμητή ιδιότητα στο G', αφού το γράφημα $G[V' \setminus K]$ δεν περιέχει ακμές, πόσο μάλλον κύκλους. Άρα, ισχύει $f(G,k) \in FVSD$. Αντίστροφα, έστω γράφημα G και $k \in \mathbb{Z}_{>0}$ τέτοια ώστε $f(G,k) \in FVSD$, τότε υπάρχει ένα $K \subseteq V$ τέτοιο ώστε να ακουμπάει κάθε κατευθυνόμενο κύκλο στο G' και $|K| \leq k$. Όμως για κάθε $\{u,v\} \in E(G)$, στο G' υπάρχει ο κατευθυνόμενος κύκλος $\{(u,v),(v,u)\}$, άρα τουλάχιστον ένα από τα v,u ανήκουν στο K, δηλαδή το K είναι κάλυμμα κορυφών του γραφήματος G, άρα $(G,k) \in VC$.

Άσκηση 4(δ) Ισχύει ότι το πρόβλημα (στο εξής θα λέγεται CSP) ανήκει στο NP. Πράγματι, ένα σύνολο ακμών ελέγχεται πολυωνυμικά για το αν είναι $s-t$ μονοπάτι και για το κατά πόσον το συνολικό κόστος διέλευσης και ο συνολικός χρόνος διέλευσης είναι μικρότερα ή ίσα από M και T αντιστοίχως. Στη συνέχεια θα δείξω ότι το CSP είναι και NP-hard, αποδεικνύοντας ότι

$$Knapsack \leq_m^p CSP$$

Έστω $((w_1,v_1),(w_2,v_2),\dots,(w_n,v_n),W,V)$ μια είσοδος του προβλήματος $Knapsack$. Δηλαδή το (w_i,v_i) είναι ζεύγος βάρους-αξίας του εκάστοτε αντικειμένου, W είναι το μέγιστο δυνατό βάρος και V είναι η ελάχιστη δυνατή αξία. Ορίζω τη συνάρτηση f ως εξής:

$$f((w_1,v_1),(w_2,v_2),\dots,(w_n,v_n),W,V) = (G,a,b,m,t,M,T)$$

όπου $G=(V,E)$ κατευθυνόμενο γράφημα, $a,b \in V$ συγκεκριμένες κορυφές του G και $m,t : E \to \mathbb{N}$ συναρτήσεις ορισμένες στις ακμές και $M,T \in \mathbb{N}$, τα οποία κατασκευάζονται ως εξής: (Χρησιμοποιώ τον συμβολισμό $[n] = \{1,2,\dots,n\}$)

$$V = \{a,b\} \cup \{x_{ij} : i,j \in [n], i \geq j\}$$

$$E = \{(a,b)\} \cup \{(a,x_{i1}) : i \in [n]\} \cup \{(x_{ij},b) : i,j \in [n], i \geq j\} \cup \{(x_{ij},x_{kl}) : i < k, j = l-1\}$$

$$m((s,t)) = \begin{cases} 0 & , \quad s = a \wedge t = b \\ w_i & , \quad s = a \wedge t = x_{i1} \\ 0 & , \quad s = x_{ij} \wedge t = b \\ w_k & , \quad s = x_{ij} \wedge t = x_{kl} \wedge i < k \wedge j = l-1 \end{cases}$$

Έστω $v = \max\{v_i : i \in [n]\}$

$$t((s,t)) = \begin{cases} nv & , \quad s = a \wedge t = b \\ iv - v_i & , \quad s = a \wedge t = v_{i1} \\ (n-i)v & , \quad s = x_{ij} \wedge t = b \\ (k-i)v - v_k & , \quad s = x_{ij} \wedge t = x_{kl} \wedge i < k \wedge j = l-1 \end{cases}$$

$$M = W \qquad\qquad T = nv - V$$

Ουσιαστικά, οι κόμβοι x_{ij} του γραφήματος κωδικοποιούν το "τοποθετώ το αντικείμενο i στο σάκο, και είναι το j-οστό αντικείμενο που τοποθετώ στο σάκο" (θεωρώ ότι τα αντικείμενα διατάσσονται σύμφωνα με το δείκτη που έχουν αρχικά και τοποθετούνται τελικά, μετά την επιλογή, με αυτή τη σειρά στο σάκο) οι κόμβοι a και b κωδικοποιούν την αρχή και το τέλος της διαδικασίας. Η συνάρτηση m εκφράζει ακριβώς το βάρος κάθε αντικειμένου που βάζω (αναφέρεται στο αντικείμενο, στο οποίο φτάνει η ακμή), ενώ η συνάρτηση t εκφράζει την ποσότητα $v - v_i$, δηλαδή θεωρώ ότι συνολικά υπάρχει αξία nv εκτός του σάκου και με το αντικείμενο i που τοποθετώ στο σάκο, αφαιρώ από αυτή την αξία ακριβώς v_i (πάλι, όταν ανατίθεται σε μια ακμή αναφέρεται στην αξία του αντικειμένου στο οποίο φτάνει η ακμή). Όπου την ποσότητα v την όρισα προηγουμένως ως το μέγιστο από τα δυνατά βάρη. (στην πραγματικότητα το v θα μπορούσε εξ' ίσου καλά να είναι οποιοσδήποτε αριθμός μεγαλύτερος ή ίσος από κάθε βάρος) Τελικά, οι ακμές κωδικοποιούν τις δυνατές κινήσεις που μπορώ να κάνω:

- Να είμαι στον κόμβο a (ο σάκος να είναι άδειος) και να πάω στον b (να επιλέξω να μη βάλω κανένα αντικείμενο). Τότε το βάρος που προσθέτω είναι 0 και η αξία που αφήνω, αφήνοντας ταυτόχρονα όλα τα αντικείμενα εκτός του σάκου είναι η μέγιστη δυνατή, δηλ. nv.
- Να είμαι στον κόμβο a και να πάω στον x_{i1} (να επιλέξω να βάλω πρώτο το αντικείμενο i). Τότε, το βάρος αυξάνεται κατά w_i και η επιπλεόν αξία που αφήνω είναι v για κάθε ένα από τα $i-1$ αντικείμενα που δεν έβαλα και ακόμη $v - v_i$ για το αντικείμενο i, δηλαδή, συνολικά $iv - v_i$.
- Να είμαι στον κόμβο x_{ij} (έχω μόλις τοποθετήσει το αντικείμενο i, με την j-οστή τοποθέτηση) και να πάω στον κόμβο b. Τότε το βαρος που προσθέτω είναι 0 και η αξία που αφήνω, αφήνοντας τα αντικείμενα $i+1, \ldots, n$ εκτός του σάκου είναι $(n-i)v$.
- Να είμαι στον κόμβο x_{ij} και να πάω στον κόμβο x_{kl}, όπου $k > i$, αφού διατρέχω τα αντικείμενα με συγκεκριμένη σειρά και $l = j+1$ αφού αναφέρομαι στην επόμενη τοποθέτηση. Τότε το βάρος που προσθέτω είναι w_k και η αξία που αφήνω είναι v για κάθε ένα από τα αντικείμενα $i+1, i+2, \ldots, k-2, k-1$, τα οποία είναι $k-i-1$ το πλήθος και ακόμη $v - v_k$ για το αντικείμενο k, δηλαδή συνολικά $(k-i)v - v_k$.

Τελικά εξετάζω αν υπάρχει μονοπάτι από το a στο b (από τον άδειο σάκο στον κλειστό σάκο), ώστε το συνολικό βάρος να μην ξεπερνάει το $M = W$ και η συνολική αξία που αφήνω εκτός του σάκου να μην ξεπερνάει το $nv - V$ (δηλαδή η αξία που θα πάρω μαζί να είναι τουλάχιστον V). Η παραπάνω συνάρτηση είναι προφανώς πολυωνυμικά υπολογίσιμη και επίσης γι αυτήν ισχύει

$$((w_1, v_1), \ldots, (w_n, v_n), W, V) \in Knapsack \Leftrightarrow f((w_1, v_1), \ldots, (w_n, v_n), W, V) \in CSP$$

Πράγματι, έστω μια είσοδος του $Knapsack$ με τους παραπάνω συμβολισμούς, που να ανήκει στο $Knapsack$. Τότε υπάρχει επιλογή δεικτών $I \subseteq [n]$ για την οποία ισχύει $\sum_{i \in I} w_i \leq W$ και $\sum_{i \in I} v_i \geq V$. Έστω $I = \{i_1 < i_2 < \cdots < i_p\}$, τότε για το μονοπάτι $P = \{(a, x_{i_1 1}), (x_{i_1 1}, x_{i_2 2}), \ldots, (x_{i_p p}, b)\}$ ισχύει ότι

$$\sum_{e \in P} m(e) = \sum_{i \in I} w_i \leq W$$

από υπόθεση και ότι

$$\begin{aligned} \textstyle\sum_{e \in P} t(e) &= i_1 v - v_{i_1} + \textstyle\sum_{f=1}^{p-1} (i_{f+1} - i_f) v - v_{i_{f+1}} + (n - i_p) v \\ &= i_1 v - v_{i_1} + v \textstyle\sum_{f=1}^{p-1} (i_{f+1} - i_f) - \textstyle\sum_{f=1}^{p-1} v_{i_{f+1}} + nv - i_p v \\ &= nv - \textstyle\sum_{f=1}^{p} v_{i_f} \leq nv - V \end{aligned}$$

από υπόθεση, επίσης. Άρα τελικά ισχύει

$$f((w_1, v_1), \ldots, (w_n, v_n), W, V) \in CSP$$

Αντίστροφα, υποθέτω μια είσοδο του $Knapsack$, τέτοια ώστε η εικόνα της μέσω της f να ανήκει στο CSP. Τότε, η επιλογή των αντικειμένων που θα βάλω στο σάκο προκύπτει άμεσα από τα i στους κόμβους x_{ij} που χρησιμοποιούνται στο μονοπάτι. Η επαλήθευση ότι οι ανισότητες ισχύουν είναι το ίδιο τετριμμένη, όπως στην αντίστροφη κατεύθυνση που εξέτασα πριν. Άρα, τελικά ισχύει ότι η αρχική είσοδος ανήκει στο $Knapsack$, το οποίο ολοκληρώνει την απόδειξη. Ακραίες περιπτώσεις όπως να μη βάλει κάποιος τίποτα στον σάκο, επίσης καλύπτονται με ακρίβεια από την παραπάνω κατασκευή.