Exam Combinatorial Optimization, 22 October 2020 (version 1)
Time: 18.45-21.30 (2 hours, 45 min.)

- It is not allowed to use any books, notes, or calculator. Just pen and paper.
- You may also obtain points for partial answers.
- Do explain each computation that you make.
- This exam has 7 questions, one on each page. The total number of points is 40

Question 1 (8 points total) The table shows an instance of the weighted Set Cover problem. A value '1' means that element e_i is included in set S_j. For example, $S_1 = \{1,2\}$. The second row gives the weights of the sets.

	S_1	S_2	S_3	S_4	S_5	S_6
	100	50	50	30	100	100
e_1	1	0	0	0	1	0
e_2	1	0	0	0	0	1
e_3	0	1	1	0	1	0
e_4	0	1	1	0	0	1
e_5	0	1	0	1	1	0
e_6	0	0	1	1	0	1

(a) (2pt.) Give an optimal solution and its value for this set cover instance.

(b) (2pt.) Give the ILP-formulation for **this** set cover instance.

(c) (2pt.) Give the the dual of the LP-relaxation of the formulation given in (b). Use a variable y_i for each element e_i.

(d) (2pt.) Apply the primal dual algorithm to this instance. Increase the dual values step by step in the order $y_1, y_2, \ldots$. Give the final y-values and the sets chosen by the algorithm.

Answer 1

(a) S_1, S_2, S_4 with value $100 + 50 + 30 = 180$.

(b)

$$
\begin{aligned}
\min \quad & 100x_1 + 50x_2 + 50x_3 + 30x_4 + 100x_5 + 100x_6 \\
s.t. \quad & x_1 + x_5 \geq 1 \\
& x_1 + x_6 \geq 1 \\
& x_2 + x_3 + x_5 \geq 1 \\
& x_2 + x_3 + x_6 \geq 1 \\
& x_2 + x_4 + x_5 \geq 1 \\
& x_3 + x_4 + x_6 \geq 1 \\
& x_1, x_2, x_3, x_4, x_5, x_6 \in \{0, 1\}.
\end{aligned}
$$

(c)

$$
\begin{aligned}
\max \quad & y_1 + y_2 + y_3 + y_4 + y_5 + y_6 \\
s.t. \quad & y_1 + y_2 \leq 100 \\
& y_3 + y_4 + y_5 \leq 50 \\
& y_3 + y_4 + y_6 \leq 50 \\
& y_5 + y_6 \leq 30 \\
& y_1 + y_3 + y_5 \leq 100 \\
& y_2 + y_4 + y_6 \leq 100 \\
& y_1, y_2, y_3, y_4, y_5, y_6 \geq 0.
\end{aligned}
$$

(d)

	y_1	y_2	y_3	y_4	y_5	y_6	
iter.0	0	0	0	0	0	0	
iter.1	100	0	0	0	0	0	S_1, S_5 tight
iter.2	100	0	0	50	0	0	S_2, S_3 tight

$y_1 = 100$, $y_4 = 50$, $y_2 = y_3 = y_5 = y_6 = 0$. Sets chosen are S_1, S_2, S_3, S_5.

Question 2 (6 points total)

$w_1 = 4, \quad C_1 = x_1 \vee x_2$
$w_2 = 4, \quad C_2 = x_2 \vee \bar{x}_3$
$w_3 = 8, \quad C_3 = x_3 \vee \bar{x}_4 \vee x_5$
$w_4 = 16, \quad C_4 = x_3 \vee \bar{x}_5 \vee x_6 \vee \bar{x}_7$
$w_5 = 16, \quad C_5 = \overline{x_2} \vee \bar{x}_3 \vee \overline{x_4} \vee \bar{x}_7$

(a) (1pt.) Shown is an instance of the weighted maximum satisfiability problem. What is the expected total weight of satisfied clauses when you assign each variable independently to true with probability 0.5? Show your computation.

(b) (1pt.) Assume that you are given an instance in which each clause has at least one positive x-variable. Argue that an optimal solution is easily found. (For example, for the given instance, only C_5 has no positive variables.)

(c) (2pt.) Argue that the following algorithm is a 0.5-approximation for the weighted maximum satisfiability problem,.

```
Consider the following two solutions and take the best:
(1) Set all variables to True.
(2) Set all variables to False.
```

(d) (2pt.) It is known that the algorithm of question (a) is a $3/4$-approximation when all clauses have at least two variables. Give an example with 2 clauses and 2 variables in each clause which shows that the approximation ratio of the algorithm of (c) is not better 0.5.

Answer 2

(a)

$$\begin{aligned} & 4(1-(\frac{1}{2})^2)+4(1-(\frac{1}{2})^2)+8(1-(\frac{1}{2})^3)+16(1-(\frac{1}{2})^4)+16(1-(\frac{1}{2})^4) \\ = \; & 4\cdot\frac{3}{4}+4\cdot\frac{3}{4}+8\cdot\frac{7}{8}+16\cdot\frac{15}{16}+16\cdot\frac{15}{16} = 3+3+7+15+15 = 43. \end{aligned}$$

(b) Set all variables to True.

(c) Every clause will be satisfied in at least one of the two solutions. Let W_1 and W_2 be the total weight of the satisfied clauses by, respectively, solution 1 and 2 and let W be the total weight of the clauses. Then, $W_1 + W_2 \geq W$ and $W \leq \text{OPT}$. Hence,

$$\max\{W_1, W_2\} \geq (W_1 + W_2)/2 \geq W/2 \geq \text{OPT}/2.$$

(d) For example, let $C_1 = x_1 \vee x_2$ and $C_2 = \bar{x}_1 \vee \bar{x}_2$ with $w_1 = w_2 = 1$. Then $\text{ALG} = 1$ and $\text{OPT} = 2$.

Question 3 (6 points total)

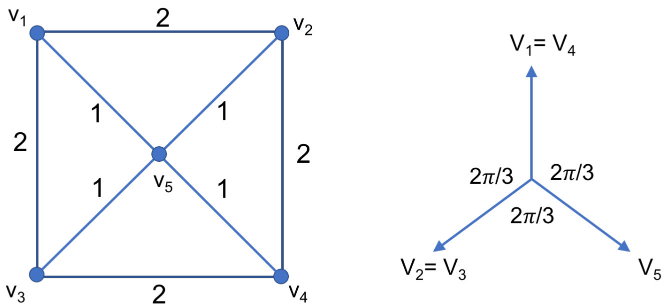
Shown below are the Quadratic Program and the Vector Program relaxation for the maximum weighted cut problem, where an instance is an edge-weighted graph $G = (V,E)$ on n vertices.

$$\text{(QP)}\quad \max \quad \frac{1}{2}\sum_{(i,j)\in E}(1 - y_i y_j)w_{ij}$$

$$s.t. \quad y_i \in \{-1, 1\} \qquad i = 1, \dots, n.$$

$$\text{(VP)}\quad \max \quad \frac{1}{2}\sum_{(i,j)\in E}(1 - v_i \cdot v_j)w_{ij}$$

$$s.t. \quad v_i \cdot v_i = 1,\ v_i \in \mathbb{R}^n \qquad i = 1, \dots, n.$$



(a) (1pt.) What is the value of the maximum weighted cut in the graph above (left)? Explain your answer.

(b) (2pt.) Shown is a solution to the VP for this instance (right). What is the VP-value of this solution?

(c) (1pt.) Argue that this is not an optimal solution to the VP.

(d) (2pt.) Remember that the max-cut algorithm of Goemans and Williamson (of chapter 6) was defined as follows: Solve VP and apply randomized rounding. What is the expected weight of the cut when you apply this algorithm? Show your computation.

Answer 3

(a) There is a cut of value 10. For example $S = \{v_1, v_4\}$. This is also the maximum cut since every triangle has at most 2 edges in the cut. Here, that means that at least two edges are not in the cut. So the maximum value is not more than 12-2=10.

(b) In this instance, the angle between any two vectors v_i, v_j with $(i, j) \in E$ is $2\pi/3$. So the value of the VP is $\frac{1}{2}(1 - -0.5)(12) = 9$.

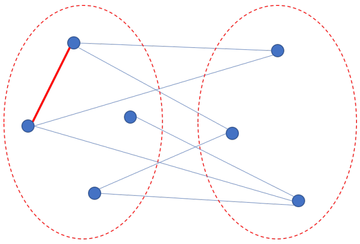
(c) The VP is a relaxation of the max cut problem so the optimal VP value is at least 10. (See a.)

(d) Each edge appears in the cut with probability $(2\pi/3)/\pi = 2/3$. So the expected total weight is $2/3 \cdot 12 = 8$.

Question 4 (6 points total) Consider the following clustering problem. Here, we want to partition the vertices of a graph into k sets (clusters) so that the total number of edges within a cluster is minimized.

k-CLUSTERING:

Instance:	Graph $G = (V, E)$
Solution:	A partition of the vertices into k sets (clusters): $S_1, S_2, \ldots, S_k \subseteq V$
Cost:	The total number of edges that have both endpoints in the same cluster.
Goal:	Minimize cost.



k=2 and the cost is 1.

(a) (2pt.) Show, by a reduction from the maximum cut problem, that Clustering is NP-hard for $k = 2$.

(b) (2pt.) Now show by a reduction from the 3-coloring problem (vertex coloring) that Clustering is NP-hard for $k = 3$.

(c) (2pt.) Now show by a reduction from the 3-coloring problem that for $k = 3$ there is no α-approximation algorithm for k-Clustering for any $\alpha \geq 1$.

Answer 4

(a) For a given instance $G = (V, E)$, let OPT_1 be the optimal value for the k-cluster problem for $k = 2$ and let OPT_2 be the optimal value for the max cut problem. Then

$$\text{OPT}_2 = |E| - \text{OPT}_1.$$

That means, we can compute the maximum cut value by computing the minimum k-cluster value. Since the, maximum cut problem is NP-hard, the k-cluster problem is NP-hard (for $k = 2$) as well.

(b) A graph is 3-colorable if and only if the optimal value for the 3-Cluster problem is 0. That means, we can decide if a graph is 3-colorable by computing the minimum 3-cluster value. Since the 3-coloring problem is NP-hard, the k-cluster problem is NP-hard (for $k = 3$) as well.

(c) Assume such an α-approximation algorithm does exist.
If the graph is 3-colorable then the optimal value for the 3-Cluster problem is 0. Then the value returned by the algorithm is at most $\alpha \cdot 0 = 0$.
If the graph is not 3-colorable then the optimal value for the 3-Cluster problem is at least 1 and so the value returned by the algorithm must also be at least 1.
That means, we can decide if a graph is 3-colorable by applying the α-approximation algorithm. Since the 3-coloring problem is NP-hard, such a (polynomial time) approximation algorithm does not exist, unless $P = NP$.

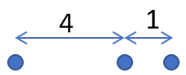
Question 5 (3 points total) Consider the k-center problem. Remember, an instance is given by a set of points in a metric space and a number k. For a given set of points in a metric space, let OPT_k be the optimal value for the k-center problem when k centers can be chosen and let OPT_{k+1} be the optimal value when $k+1$ centers can be chosen.

(a) (2pt.) Give (draw) an example (a set of points and distances) for which $\mathrm{OPT}_1 > 3\mathrm{OPT}_2$.

(b) (1pt.) Argue that for any k and $\alpha > 1$ there exists an instance for which $\mathrm{OPT}_k/\mathrm{OPT}_{k+1} > \alpha$.

Answer 5

(a) See figure. $\mathrm{OPT}_1 = 4$ and $\mathrm{OPT}_2 = 1$.



(b) See figure. $\mathrm{OPT}_k = A$ and $\mathrm{OPT}_{k+1} = 1$. Take $A > \alpha$.

NB. Many more answers are possible. For example:

(a) Take 2 points at arbitrary distance. Then $\mathrm{OPT}_1 > 0$ and $\mathrm{OPT}_2 = 0$. Then $\mathrm{OPT}_1/\mathrm{OPT}_2 = \infty$.

(b) Take $k+1$ points at at arbitrary distance. Then $\mathrm{OPT}_k > 0$ and $\mathrm{OPT}_{k+1} = 0$. Then $\mathrm{OPT}_k/\mathrm{OPT}_{k+1} = \infty$.

Question 6 (4 points total) Below is the ILP for the Uncapacitated Facility Location (UFL) problem.

$$\begin{aligned}
\text{(ILP)} \quad \min \quad & Z = \sum_{i \in F} f_i y_i + \sum_{i \in F, j \in D} c_{ij} x_{ij} \\
s.t. \quad & \sum_{i \in F} x_{ij} = 1 && \text{for all } j \in D, \\
& x_{ij} \leqslant y_i && \text{for all } i \in F, j \in D, \\
& x_{ij} \in \{0,1\} && \text{for all } i \in F, j \in D, \\
& y_i \in \{0,1\} && \text{for all } i \in F.
\end{aligned}$$

Here, F is the set of facilities and D is the set of clients,

We make the following change to the UFL: *In stead of connecting all clients we only need to connect* 3 *clients and we need to find those 3 for which the total cost (connection plus opening) is minimized.*

(a) (3pt.) Give an ILP for this problem based on the ILP above.

(b) (1pt.) Argue that an optimal solution can be found in polynomial time.

Answer 6

(a) Add the following constraint

$$\sum_{j \in D} \sum_{i \in F} x_{ij} = 3$$

and replace the fist equality by an inequality:

$$\sum_{i \in F} x_{ij} \leq 1 \text{ for all } j \in D.$$

(b) If only 3 clients get connected then a most 3 facilities are opened. So we can enumerate over all possible solutions in polynomial time. We need to check at most $O(|D|^3 |F|^3)$ solutions.

Question 7 (7 points total)

For this question, we first recap the theory of the Prize-collecting Steiner Tree problem (as written in the lecture notes).

PRIZE-COLLECTING STEINER TREE:

Instance:	$G = (V,E)$ and a cost c_e for every edge $e \in E$ and a penalty π_i for every vertex $i \in V$. Also given is a root $r \in V$.
Solution:	Tree T containing r. Let $V(T)$ be the vertices in T
Value:	$\sum_{e\in T} c_e + \sum_{i\in V-V(T)} \pi_i$.
Goal:	Find a solution of minimum cost.

In words, the goal is to find a tree containing r that minimizes the cost of all edges in the tree plus the penalty cost for the vertices that are not connected by T. An ILP for the problem is as follows.

$$
\begin{aligned}
\text{(ILP)} \quad \min \quad & Z = \sum_{e\in E} c_e x_e + \sum_{i\in V} \pi_i (1 - y_i) \\
s.t. \quad & \sum_{e\in\delta(S)} x_e \geqslant y_i, && \text{for all } i,S \text{ with } i \in S \subseteq V - r, \\
& y_r = 1, \\
& x_e \in \{0,1\}, && \text{for all } e \in E, \\
& y_i \in \{0,1\}, && \text{for all } i \in V.
\end{aligned}
$$

Algorithm : $(0 < \alpha < 1)$
Step 1: Solve the LP-relaxation of (ILP) $\rightarrow \quad x^*, y^*, Z^*_{LP}$.
Step 2: Let $U = \{i \mid y^*_i \geqslant \alpha\}$. Construct a Steiner tree T on U.

Lemma 1 *The connection cost for the Steiner tree T is*

$$\sum_{e\in T} c_e \leqslant \frac{2}{\alpha} \sum_{e\in E} c_e x^*_e.$$

Lemma 2 *The total penalty cost is*

$$\sum_{i\in V-V(T)} \pi_i \leqslant \frac{1}{1-\alpha} \sum_{i\in V} (1 - y^*_i)\pi_i.$$

Theorem 1 *The algorithm above with $\alpha = 2/3$ is a 3-approximation algorithm for the Prize-collecting Steiner Tree problem.*

Proof: The sum of connection and penalty cost is at most

$$\frac{2}{\alpha} \sum_{e\in E} c_e x^*_e + \frac{1}{1-\alpha} \sum_{i\in V} (1 - y^*_i)\pi_i \leqslant \max\left\{\frac{2}{\alpha}, \frac{1}{1-\alpha}\right\} Z^*_{LP}.$$

The value $\max\left\{\frac{2}{\alpha}, \frac{1}{1-\alpha}\right\}$ is 3 for $\alpha = 2/3$. Hence, the total cost of the solution of the algorithm is at most

$$3Z^*_{LP} \leqslant 3\text{OPT}.$$

(a) (2pt.) If we would apply this algorithm with $\alpha = 1/2$, then what approximation ratio do you get? Show your computation. You may refer to the theory above.

(b) (1pt.) In the *unrooted* version of the problem, no root r is given. That means, any tree is allowed. Show that if you have a β-approximation algorithm for the rooted version (for example with $\beta = 3$), then you can easily get a β-approximation for the unrooted version.

Now, we make the following adjustment to the Prize-collecting Steiner Tree problem:
In stead of one root r we are given two roots, r_1 and r_2. The problem is to find two trees T_1, T_2 with $r_1 \in T_1$ and $r_2 \in T_2$. The goal is to minimizes the cost of all edges in the two trees plus the penalty cost for the vertices that are not in any tree.

(c) (2pt.) For some given instance, let $\text{OPT}(r_1, r_2)$ be the optimal value for the Prize-collecting Steiner Tree problem with roots r_1 and r_2 and let $\text{OPT}(r_1)$ be the optimal value for the standard Prize-collecting Steiner Tree problem with one root r_1. Argue that $\text{OPT}(r_1, r_2) \leq \text{OPT}(r_1)$.

(c) (2pt.) Write down an ILP for this problem based on the ILP above. Use variables x_e^1 and x_e^2 and y_i^1 and y_i^2.

Answer 7

(a) The analysis works for any α with $0 < \alpha < 1$. Hence, the ratio is $\max\left\{\frac{2}{\alpha}, \frac{1}{1-\alpha}\right\} = \max\{4, 2\} = 4$.

(b) Apply the algorithm for the rooted version for all $v \in V$ as the root and take the best solution. Explanation: Assume that in the unrooted version the optimal solution has root v_0. Then when we apply the rooted algorithm with root v_0, the value found is at most β times the optimal unrooted value.

(c) Let $T(r_1)$ and $T(r_1, r_2)$ be the optimal trees. We consider 2 cases.

If $r_2 \in T(r_1)$ then deleting any edge on the path in the tree between r_1 and r_2 gives a solution for the 2-rooted problem of at most the same cost. So in this case, $\text{OPT}(r_1, r_2) \leq \text{OPT}(r_1)$.

If $r_2 \notin T(r_1)$ then the cost of this solution in the 2-rooted problem is not larger since we do not pay the penalty for r_2 then. So also in this case, $\text{OPT}(r_1, r_2) \leq \text{OPT}(r_1)$.

(d)

$$
\begin{aligned}
\text{(ILP)} \quad \min \quad & Z = \sum_{e \in E} c_e (x_e^1 + x_e^2) + \sum_{i \in V} \pi_i (1 - y_i^1 - y_i^2) & \\
s.t. \quad & \sum_{e \in \delta(S)} x_e^1 \geqslant y_i^1, & \text{for all } i, S \text{ with } i \in S \subseteq V - r_1, \\
& \sum_{e \in \delta(S)} x_e^2 \geqslant y_i^2, & \text{for all } i, S \text{ with } i \in S \subseteq V - r_2, \\
& y_i^1 + y_i^2 \leq 1 & \text{for all } i \in V, \\
& y_{r_1} = 1, & \\
& y_{r_2} = 1, & \\
& x_e^1, x_e^2, \in \{0, 1\}, & \text{for all } e \in E, \\
& y_i^1, y_i^2 \in \{0, 1\}, & \text{for all } i \in V.
\end{aligned}
$$

Explanation: For each tree, we have a cut constraint just like the single root version. For the penalty cost we need the additional constraint that for each vertex, at most one of the y-values is 1. With this constraint, a similar constraint for the edges $(x_e^1 + x_e^2 \leq 1)$ is not needed since it will always hold in an optimal solution. It is OK to add it though.