ORIE 6300 Mathematical Programming I November 21, 2008

Problem Set 11

Due Date: December 5, 2008

1. (20 points) In the *k-center problem*, we are given a set of locations V and distances d_{ij} between all locations $i, j \in V$. Distances are symmetric ($d_{ij} = d_{ji}$ for all $i, j \in V$) and obey the triangle inequality ($d_{ik} \leq d_{ij} + d_{jk}$ for all $i, j, k \in V$). We are also given an input k, which is an integer. The goal of the problem is to choose a subset $S \subseteq V$ of k *centers* to minimize the distance of every location to the nearest center; that is, we wish to minimize $\max_{i \in V} \min_{j \in S} d_{ij}$. Let S^* be an optimal solution to the problem, and let $OPT = \max_{i \in V} \min_{j \in S^*} d_{ij}$.

   (a) (10 points) Consider an algorithm for this problem that starts by picking some center arbitrarily from V, then repeatedly picks the next center to be as far away as possible from the previously chosen centers. That is, if S is the set of currently chosen centers, we pick the next center to be the location i that maximizes $\max_{i \in V} \min_{j \in S} d_{ij}$. Show that this algorithm always returns a solution S such that $\max_{i \in V} \min_{j \in S} d_{ij} \leq 2OPT$.

   (b) (10 points) Suppose that there exists some polynomial-time algorithm that always finds a solution S' to the k-center problem with $\max_{i \in V} \min_{j \in S'} d_{ij} \leq \rho \times OPT$, where $\rho < 2$. Prove that such an algorithm would imply that $P = NP$. (Hint: consider the problems shown to be NP-complete in class and on previous problem sets, and show that such an algorithm would imply a polynomial-time algorithm for an NP-complete problem.)

2. (25 points) Recall from the midterm the *maximum multicommodity flow problem*. In this problem we are given a directed graph G with nodes V and directed arcs A, and k source-sink pairs (s_i, t_i), where $s_i, t_i \in V$ for $i = 1, \ldots, k$. We may send flow only from a source s_i to the corresponding sink t_i. The goal is to send as much flow as possible from the sources s_i to their corresponding sinks t_i. Each arc $a \in A$ has a capacity u_a; we may not send more than u_a total units of flow through arc a.

   We can write the problem as a linear program. Let $\mathcal{P}_i$ be the set of paths in G from s_i to t_i. Our LP will have a variable x_P for each $P \in \mathcal{P}_i$ for each $i = 1, \ldots, k$. Then the maximum multicommodity flow problem can be modelled as the following linear program:

$$
\begin{aligned}
\text{Max} \quad & \sum_{i=1}^{k} \sum_{P \in \mathcal{P}_i} x_P \\
\sum_{i=1}^{k} \sum_{P \in \mathcal{P}_i : a \in P} x_P & \leq u_a \qquad \forall a \in A \\
x_P & \geq 0 \forall i = 1, \ldots k, \forall P \in \mathcal{P}_i.
\end{aligned}
$$

   The dual of the maximum multicommodity flow linear program is

$$
\begin{aligned}
\text{Min} \quad & \sum_{a \in A} u_a \ell_a \\
\sum_{a \in P} \ell_a & \geq 1 \qquad \forall P \in \mathcal{P}_i, \forall i = 1, \ldots, k \\
\ell_a & \geq 0 \qquad \forall a \in A.
\end{aligned}
$$

(a) (10 points) Explain how to solve the dual in polynomial time (given the maximum multicommodity flow instance as input) using the ellipsoid method.

(b) (15 points) Unlike the simplex method, the ellipsoid method doesn't automatically give an optimal solution to the dual of the LP being solved. Explain how to solve the original maximum multicommodity flow LP given the execution of the ellipsoid method in solving the dual above.

3. (25 points) Recall the assignment problem, and suppose that we draw the costs c_{ij} uniformly from $[0, 1]$ for each i, j, where $i = 1, \ldots, n$ and $j = 1, \ldots, n$. In this problem, we show that the expected optimal solution to the assignment problem has value at most 2 (i.e. it doesn't depend on n).

   (a) (3 points) Let C be the random variable such that $C = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}$. Compute the expected value of C.

   (b) (15 points) We know from the recitation on October 1 that any basis is given by a tree T in the bipartite graph of edges, and that we can choose one of the dual variables (say u_n) and set it to zero arbitrarily. Let X_T be the event that T is the optimal basis. We now wish to compute $E[C|X_T]$. To do this, we see that the values of the dual variables u_i and v_j are fixed by the realizations of the values of the c_{ij} for all $(i, j) \in T$, but that for any $(i, j) \notin T$, $c_{ij} \geq u_i + v_j$, and this changes the distribution and thus the expected value of c_{ij} for these non-tree edges. Use this to show that

   $$E[C|X_T] \geq \frac{1}{2} \sum_{(i,j) \in T} (u_i + v_j) + \sum_{(i,j) \notin T} \left(\frac{1}{2} + \frac{1}{2}(u_i + v_j) \right).$$

   (c) (7 points) We know that $E[C] = \sum_T E[C|X_T] \Pr[X_T]$. Ignoring the possibility that there may be more than one optimal basis for a given set of costs, and letting Z^* be a random variable denoting the cost of an optimal assignment, use the above to prove that $E[Z^*] < 2$.