Solutions for assignment 5 of COMP 360

TA:Xing Shi Cai (newptcai@gmail.com)

December 7, 2012

Question 1 (50 points)

Proof. Let t_j be the processing time of the j-th job. Let T' be the optimal *makespan*. There are two trivial lower bounds for T':

$$
\begin{aligned}
T' &\geq \frac{\sum_j t_j}{k} \\
T' &\geq \max_j t_j.
\end{aligned}
$$

Assume that when the greedy algorithm terminates the ℓ-th machine has the highest workload, which we denote by T_ℓ. Let t_r be the last job added to machine ℓ. Since the algorithm only adds a job to a machine when it has the minimal workload, $T_\ell - t_r$ is at most the average load before t_r is added, which is at most

$$
\frac{\left(\sum_j t_j\right) - t_r}{k}.
$$

Put differently, we have

$$
T_\ell - t_r \leq \frac{\left(\sum_j t_j\right) - t_r}{k}.
$$

Plugging in the two inequalities given at the beginning, we have

$$
\begin{aligned}
T_\ell &\leq t_r + \frac{\left(\sum_j t_j\right) - t_r}{k} \\
&= \frac{\sum_j t_j}{k} + t_r \left(1 - \frac{1}{k}\right) \\
&\leq T' + T' \left(1 - \frac{1}{k}\right) \\
&= \frac{2k-1}{k} T'.
\end{aligned}
$$

□

Question 2 (50 points)

Proof. The first part of the algorithm is the same as the greedy algorithm given in course textbook for selecting k centers. Let $\{v_1, \ldots, v_k\}$ denote the points selected as centers. The second part of algorithm is just partition all the points into k disjoint sets $\mathcal{D}_1, \ldots, \mathcal{D}_k$ such that

$$\mathcal{D}_i = \{x \mid \operatorname{dist}(x, v_i) = \min_{1 \le j \le k} \operatorname{dist}(x, v_i)\}.$$

Put it differently, we partition the points into k subsets such that each points and their closest center are in the same subset. Let $\mathcal{D}_r$ be the subset which contains two point u and v such that

$$\operatorname{dist}(u, v) = \max_{1 \le j \le k} \max_{x, y \in \mathcal{D}_j} \operatorname{dist}(x, y),$$

i.e., $\operatorname{dist}(u, v)$ is the output of the greedy algorithm. Let v_r be the center of $\mathcal{D}_r$. It follows from the triangle inequality that

$$\operatorname{dist}(u, v) \le \operatorname{dist}(u, v_r) + \operatorname{dist}(v, v_r).$$

Without losing of generality, we can assume that $\operatorname{dist}(u, v_r) \ge \operatorname{dist}(v, v_r)$. Writing $d = \operatorname{dist}(u, v_r)$, by the last inequality, we have

$$\operatorname{dist}(u, v) \le 2d. \tag{1}$$

First note that, since v_r is the closest center to u, we must have

$$\operatorname{dist}(u, v_i) \ge d,$$

for all $1 \le i \le k$. Also note that, the distance between any two centers must be at least d, otherwise a center would have less distance to S as u does. Therefore, we have in total $k + 1$ points whose pairwise distance are all at least d, which means there is no way we can partition all the points into k subsets such that within each subset the maximal distance between points are less than d. Put differently, we have $d' \ge d$, where d' is the optimum of this partition problem. It then follows from (1) that

$$\operatorname{dist}(u, v) \le 2d \le 2d'.$$

In other words, the algorithm is a 2-factor approximation. □

Question 3 (50 points)

Notations: Let $E(A, B)$ denote the number of edges from A to B. Let E_{opt} be the maximal possible value of $E(A, B)$.

(a). For each vertex u, choose $u \in A$ with probability $1/2$. Then for each edge $(u, v) \in E$, $\mathbb{P}\{[u \in A] \cap [v \in B]\} = 1/4$. Taking expectation of $E(A, B)$, we have

$$\mathbb{E}[E(A, B)] = \frac{1}{4}|E| \ge \frac{1}{4} E_{\text{opt}}.$$

where $|E|$ is the total number of edges and is a trivial upper bound of E_{opt}. □

(b). Let x_u and z_{uv} be a solution of the ILP. Let $u \in A$ if $x_u = 1$ and let $u \in B$ otherwise. The conditions that $z_{uv} \le x_u$ and $z_{uv} \le 1 - x_v$ guarantee that z_{uv} can only take positive value when $u \in A$ and $v \in B$. Since the ILP maximizes $\sum_{(u,v) \in E} z_{uv}$, z_{uv} would take the maximal value 1 when $u \in A$ and $v \in B$. Thus, $E(A, B) = \sum_{(u,v) \in E} z_{uv}$. Therefore, the optimal solution of the ILP gives us E_{opt}. □

(c). Let z^*_{uv} and x^*_u be an optimal solution for the LP. Let $z'_{uv} = 1$ if $u \in A$ and $v \in B$ after the random rounding, and let $z'_{uv} = 0$ otherwise. Thus $\sum_{(u,v)\in E} z'_{uv}$ is the number of edges from A to B. Taking expectation, we have

$$\begin{aligned}
\mathbb{E}\left[z'_{uv}\right] &= \mathbb{P}\left\{z'_{uv} = 1\right\} \\
&= \mathbb{P}\left\{[u \in A] \cap [v \in B]\right\} \\
&= \mathbb{P}\left\{u \in A\right\} \mathbb{P}\left\{v \in B\right\} \\
&= \left(\frac{1}{4} + \frac{x^*_u}{2}\right)\left(\frac{3}{4} - \frac{x^*_v}{2}\right) \\
&\geq \left(\frac{1}{4} + \frac{z^*_{uv}}{2}\right)^2 \\
&= \left(\frac{1}{4} - \frac{z^*_{uv}}{2}\right)^2 + \frac{z^*_{uv}}{2} \\
&\geq \frac{z^*_{uv}}{2}.
\end{aligned}$$

Therefore, taking expectation of the number of edges from A to B, we have

$$\begin{aligned}
\mathbb{E}\left[E(A,B)\right] &= \mathbb{E}\left[\sum_{(u,v)\in E} z'_{uv}\right] \\
&= \sum_{(u,v)\in E} \mathbb{E}\left[z'_{uv}\right] \\
&\geq \sum_{(u,v)\in E} \frac{z^*_{uv}}{2} \\
&= \frac{1}{2}\operatorname{opt}(\mathrm{LP}) \\
&\geq \frac{1}{2}E_{\mathrm{opt}}. \qquad \square
\end{aligned}$$