1 Approximation Algorithm II Homework 1

2 Approximation Algorithm II Homework 2

3 Approximation Algorithm II Homework 3

A Primal-dual algorithm for K-median
Symbols:

- $G = (F \cup C, E)$ bipartite graph, where F is the center set and C the client set
- $c : E \to \mathbb{R}_+$ metric function

Goal, find $S \subset F$ having at most k elements. Such that

$$\sum_{v \in C} \min_{f \in S} c(v, f)$$

This problem is similar to the facility location problem, except that there is no cost for opening a facility. We want to design a primal-dual algorithm achieving a 6-approximation based on the primal dual algorithm seen during the lectures.

The standard LP formulation, LP1, is as follows:
Form of LP1:

$$\begin{array}{ll}
\min & \displaystyle\sum_{j \in C} \min_{i \in F} x_{ij} \cdot c((i, j)) \\
\text{subject to} & \\
\forall j \in C & \displaystyle\sum_{i \in F} x_{ij} \geq 1 \\
\forall j \in C, i \in F & y_i - x_{ij} \geq 0 \\
 & \displaystyle\sum_{j \in C} y_j \leq k \\
\forall j \in C, i \in F & x_{ij} \geq 0 \\
\forall i \in F & y_i \geq 0
\end{array}$$

In order to have a LP, LP2, that is closer to the one we had for facility location, we will consider a slightly different LP.

For a given $\lambda \geq 0$, we define the following LP.
Form of LP2:

$$\begin{array}{lr}
\min & \displaystyle\sum_{j \in C} \min_{i \in F} x_{ij} \cdot c\left((i,j)\right) + \lambda \left(\sum_{i \in F} y_i \right) - \lambda k \\
\text{subject to} & \\
\forall j \in C & \displaystyle\sum_{i \in F} x_{ij} \geq 1 \\
\forall j \in C, i \in F & y_i - x_{ij} \geq 0 \\
\forall j \in C, i \in F & x_{ij} \geq 0 \\
\forall i \in F & y_i \geq 0
\end{array}$$

The parameter λ is a way to penalize the LP for opening a lot of facilities. Indeed, we would like to find a good value for λ so that an optimal solution for the LP will open at most k facilities.

Observe that any feasible solution for the linear programming relaxation of the k-median problem, LP1, is also feasible for LP2.

Moreover, any feasible solution for LP2 is also a solution for LP1, assuming $\lambda \geq 0$.

My Remark: The above sentence is not right! only if there is a $\max_\lambda$ before the term in objective! Or if here feasible solution means optimal, then it's right.

Problem 3.1 *Give the dual of LP2, DUAL2.*

Solution: Let α_j, $j \in C$ be the dual of first constraints, and β_{ij}, $i \in F$, $j \in C$ be the dual of second constraints, we have following:

$$\begin{array}{lr}
\max & \displaystyle\sum_{j \in C} \alpha_j - \lambda k \\
\text{subject to} & \\
\forall j \in C,\ i \in F & \alpha_j - \beta_{ij} \leq c\left((i,j)\right) \\
\forall i \in F & \displaystyle\sum_{j \in C} \beta_{ij} \leq \lambda \\
\forall i \in F,\ j \in C & \beta_{ij} \geq 0 \\
\forall j \in C & \alpha_j \geq 0
\end{array}$$

□

We now would like the primal-dual algorithm seen during the lectures (the one that achieves a 3-approximation) in which all the facility cost are equal to some $\lambda \geq 0$.

Recall that in Lecture 8, we have seen that

$$\sum_{\text{cluster}C_0} f_{i_C} + \sum_{j \in C_0} c\left((i,j)\right) \leq 3 \sum_{j \in C} \alpha_j$$

In fact, it is possible to show that

$$\sum_{\text{cluster}C_0} 3 \cdot f_{i_C} + \sum_{j \in C_0} c\left((i,j)\right) \leq 3 \sum_{j \in C} \alpha_j$$

Problem 3.2 *Suppose that the f_i are all equal to λ and that the algorithm opens a set S of facilities, what can you deduce from the above formula?*

Solution: Substitute λ into f_i, we have

$$\sum_{\text{cluster}C_0} 3\lambda + \sum_{j \in C_0} c\left((i,j)\right) \leq 3 \sum_{j \in C} \alpha_j$$

Note $\sum_{\text{cluster}C_0} \lambda = \sharp\text{cluster} \cdot \lambda$,

$$3\sharp\text{cluster} \cdot \lambda + \sum_{\text{cluster}C_0} \sum_{j \in C_0} c\left((i,j)\right) \leq 3 \sum_{j \in C} \alpha_j$$

Minus $3k\lambda$ on both side,

$$3(\sharp\text{cluster} - k) \cdot \lambda + \sum_{\text{cluster}C_0} \sum_{j \in C_0} c\left((i,j)\right) \leq 3 \sum_{j \in C} \alpha_j - 3k\lambda$$

That is, the cost for LP2 is 3-approximation optimal

$$3(\sharp\text{cluster} - k) \cdot \lambda + \sum_{\text{cluster}C_0} \sum_{j \in C_0} c\left((i,j)\right) \leq 3OPT$$

□

Problem 3.3 *Suppose further that the set S contains exactly k facilities, what can you deduce from the cost of the solution output by the algorithm and the value of the dual?*

Solution: Note $\sharp$cluster $= |S|$, we then have cost for LP1 is 3-approximation. That is

$$cost(S) \leq 3 \cdot \left(\sum_{j \in C} \alpha_j - |S| \cdot \lambda\right)$$

$\square$

We now want to find the value of λ that would lead the algorithm to open exactly k facilities.

In order to do so, we will do a bisection search and maintain a lower bound λ_1 and an upper bound λ_2 on the value of the optimal λ.

We start with $\lambda_1 = 0$ and $\lambda_2 = \sum_{j \in C} \sum_{i \in F} c\left((i, j)\right)$.

Problem 3.4 *Prove that in the case of $\lambda = \lambda_1$, the algorithm opens at least k facilities and in the case of $\lambda = \lambda_2$ the algorithm opens less than k facilities.*

Solution: When $\lambda = 0$, there are no cost of openning more facilities than k. Hence as much as possible facilities will be opened.

When $\lambda = \lambda_2$, opening a new facility more than k will induce a cost more than total sum of service cost, hence it will not be opened, because we can always serve a customer using a path which cost less than the cost of openning a new facility. $\square$

We start by running the algorithm for λ_1, which returns a set of facilities S_1. Then we run the algorithm for λ_2 and obtain a set of facilities S_2. If $|S_1| > k$ and $|S_2| < k$, we run the algorithm on the value $\lambda = (\lambda_1 + \lambda_2)/2$.

The algorithm outputs a set S of facilities. If $|S| > k$, we set $\lambda_1 = (\lambda_1 + \lambda_2)/2$, $S_1 = S$ and repeat.

Otherwise we set $\lambda_2 = (\lambda_1 + \lambda_2)/2$, $S_2 = S$ and repeat.

We repeat until we obtain a set S of k facilities or $\lambda_2 - \lambda_1$ is small enough.

In the later case, we explain how to combine S_1 and S_2 to obtain a solution with k facilities.

Let $c_{\min}$ be the smallest assignment cost greater than 0. We run the bisection search until we get a set S of k facilities or until $\lambda_2 - \lambda_1 \le \epsilon c_{\min}/(3|F|)$, for some fixed $\epsilon > 0$.

If we have not terminated with a solution with exactly k facilities, the algorithm terminates with solutions S_1 and S_2 and (corresponding) dual solutions (α^1, β^1) and (α^2, β^2) such that $|S_1| > k > |S_2|$.

Problem 3.5 *Use Question 3 to derive an inequality connecting the value of the solution induced by the set S_1, cost($S1$), and (α^1, β^1) and λ_1. Derive a similar bound for the value of the solution induced by S_2, cost(S_2).*

Solution: From Q3, we have

$$cost(S_1) \leq 3 \cdot (\sum_{j \in C} \alpha_j^1 - |S_1| \cdot \lambda_1)$$

and

$$cost(S_2) \leq 3 \cdot (\sum_{j \in C} \alpha_j^2 - |S_2| \cdot \lambda_2)$$

□

Without loss of generality, we can assume that $0 < c_{\min} \leq OPT$, since if $OPT = 0$ then it is easy to compute an optimal solution.

We now pick $\delta_1, \delta_2 > 0$ such that $\delta_1 + \delta_2 = 1$ and $\delta_1|S_1| + \delta_2|S_2| = k$.

We can then get a dual solution $(\tilde{\alpha}, \tilde{\beta})$ by letting $\tilde{\alpha} = \delta_1\alpha_1 + \delta_2\alpha_2$ and $\tilde{\beta} = \delta_1\beta_1 + \delta_2\beta_2$.

Note that $(\tilde{\alpha}, \tilde{\beta})$ is feasible for the DUAL2 with facility costs λ_2 since it is a convex combination of two feasible dual solutions. We can now prove the following lemma.

It states that the convex combination of the costs of the solutions induced by S_1 and S_2 must be close to the cost of an optimal solution.

Lemma 3.1

$$\delta_1 cost(S_1) + \delta_2 cost(S_2) \leq (3 + \delta_1\epsilon)OPT$$

Problem 3.6 *Using Q5, prove*

$$cost(S_1) \leq 3(\sum_{j \in C} \alpha_j^1 - \lambda_2|S_1|) + \epsilon OPT$$

Solution: Using fact:

$$\lambda_2 - \lambda_1 \leq \epsilon c_{\min}/(3|F|)$$

which is

$$-\lambda_1 \leq \epsilon c_{\min}/(3|F|) - \lambda_2$$

Times $3|S_1|$ on both side,

$$-3\lambda_1|S_1| \leq 3|S_1| \cdot \epsilon \frac{c_{\min}}{3|F|} - 3|S_1| \cdot \lambda_2$$

Using fact $|S_i| \leq |F|$ and $c_{\min} \leq OPT$, we have

$$-3\lambda_1|S_1| \leq \epsilon \cdot OPT - 3|S_1| \cdot \lambda_2$$

Substitute to Q5, we have

$$cost(S_1) \leq 3\sum_{j\in C}\alpha_j^1 - 3\lambda_2|S_1| + \epsilon OPT$$

Which is what we want. $\square$

Problem 3.7 *Using Q6 and convex combination of $cost(S_2)$ to prove Lemma 3.1.*

Solution: Combination of LHS gives

$$\begin{aligned}
&3\sum_{j\in C}(\delta_1\alpha_j^1 + \delta_2\alpha_j^2) - 3\lambda_2(\delta_1|S_1| + \delta_2|S_2|) + \delta_1\epsilon OPT \\
=&3\sum_{j\in C}\tilde{\alpha_j} - 3\lambda_2 \cdot k + \delta_1\epsilon OPT \\
\leq&3 \cdot OPT + \delta_1\epsilon OPT
\end{aligned}$$

The second inequality is because of $3\sum_{j\in C}\tilde{\alpha_j} - 3\lambda_2 \cdot k \leq OPT(DUAL2) \leq OPT$. $\square$

Discuss for δ_1 greater or equal to 0.5 or opposite.

In the case $\delta_2 \geq 1/2$, we return S_2. (Note $|S_2| < k$ and S_2 is feasible)

Problem 3.8 *Using Lemma1 and $\delta_2 \geq 1/2$, show S_2 is a solution of cost at most $2(3+\epsilon)OPT$.*

Solution: From Lemma1 we have

$$\delta_2 cost(S_2) \leq (3 + \delta_1\epsilon)OPT \leq (3+\epsilon)OPT$$

Using $\delta_2 \geq 1/2$, we have

$$cost(S_2) \leq 2(3+\epsilon)OPT$$

. $\square$

Now we assume $\delta_2 < 1/2$.

For each $i \in S_2$ open some closest $h \in S_1$ so that total opened facilites are of number $|S_2|$. Because $|S_1| > k > |S_2|$, we can choose randomly $k - |S_2|$ of $|S_1| - |S_2|$ remaining unopened set, and open them. Let S be the resulting set of all opend facilities. (So there are exactly k openned facilities)

My Remark: there might be a big overlap between S_1 and S_2, it doesn't matter. If $i \in S_1 \cap S_2$, then the closest one is just i itself, hence all $S_1 \cap S_2$ is opened first.

We show the expected cost of S is at most $2(3 + \epsilon)OPT$.

We give a bound on the expected cost of assigning a given client j to a facility opened by the randomized algorithm.

Let us suppose that the facility $f_1 \in S_1$ is the open facility in S_1 closest to j. This means that $c((f_1, j))$ is the contribution of client j to the cost of S_1, c_j^1.

My remark: the notation here means $c_j^1 = c(f_1, j)$ and not any other $c(i', j)$ because in the primal setting , there is a minimum there.

Let $f_2 \in S_2$ be the open facility in S_2 that is the closest to j. Again, $c((f_2, j))$ is the contribution of client j to the cost of S_2, c_j^2.

Recall that $\frac{k-|S_2|}{|S_1|-|S_2|} = \delta_1$. (because δ_i is chosen as such convex combination)

Problem 3.9 *What is the probability that the randomized algorithm opens f_1?*

Solution: Because f_1 is the nearest to j in S_1, if it's not open due to far from any of S_2, then j should also be far from all S_2. Hence the probability is δ_1. □

If f_1 is open, then j is assigned to f_1. Otherwise, we assign j to the closest facility of S_1 closest to f_2. Let i be this facility. Recall that by the triangle inequality we have that $c((i, j)) \leq c((j, f_2)) + c((f_2, i))$.

Problem 3.10 *Give an upper bound on the distance of $c((i, f_2))$ based on the distance from $c((f_1, f_2))$.*

Solution: $c(i, f_2) \leq c(f_1, f_2)$ or it's not closest. □

Problem 3.11 *Show that $c((i, j)) \leq c_j^1 + 2c_j^2$.*

Solution:

$$c(i, j) \leq c(i, f_2) + c(f_2, j)$$

And

$$c(i, f_2) \leq c(f_1, f_2)$$

Hence

$$c(i, j) \leq c(f_1, f_2) + c(f_2, j)$$

Note

$$c(f_1, f_2) \leq c(f_1, j) + c(f_2, j)$$

We have,

$$c(i, j) \leq c(f_1, j) + 2c(f_2, j) = c_j^1 + 2c_j^2$$

. $\square$

Problem 3.12 *Based on Questions 9 and 11, show that the expected cost for client j is at most $\delta_1 c_j^1 + \delta_2(c_j^1 + 2c_j^2)$.*

Solution: Because the i is the random variable. The probability: $P[c(i,j) \leq c_j^1] \geq \delta_1$. However using the following:

$$E[c(i,j)] \leq P[c(i,j) \leq c_j^1]c_j^1 + (1 - P[c(i,j) \leq c_j^1])(c_j^1 + 2c_j^2)$$

We have

$$E[c(i,j)] \leq \delta_1 c_j^1 + \delta_2(c_j^1 + 2c_j^2)$$

. $\square$

Problem 3.13 *Conclude the proof of this case using Question 12 and the fact that $\delta_2 < 1/2$.*

Solution: Take expectation, we have:

$$E[cost(S)] \leq \delta_1 cost(S_1) + \delta_2(cost(S_1) + cost(S_2))$$

Then using lemma1 and $\delta_2 cost(S_1) \leq \delta_1 cost(S_1) + \delta_2 cost(S_2)$, we have

$$E[cost(S)] \leq 2(3 + \epsilon)OPT$$

.

$\square$