IOE 691: Approximation Algorithms	Date: 02/13/2017

Lecture Notes: Metric Facility Location (LP Rounding).

Instructor: Viswanath Nagarajan Scribe: Runsheng Du

1 Linear Programming

In general, linear programming (LP) can be expressed as:

$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to:} \quad & Ax \leq b \\ & x \geq 0, \end{aligned}$$

where x is a vector of n variables, c is the linear objective, A is an $m \times n$ matrix and b is an m dimensional vector.

Theorem 1.1. *Any linear program in n variables and m constraints can be solved optimally in $poly(n, m, \log D)$ time, where D is the largest entry in c, A, b.*

There are two methods to solve linear programs in polynomial time: ellipsoid and interior point. In most practical instances, the simplex method does better; although its runtime is not polynomial.

An integer program is similar to an LP, except that variables may be restricted to integer values. IPs are NP-hard to solve. However many approximation algorithms work by (1) formulating an integer program, (2) relaxing integrality to obtain an LP relaxation which can be solved efficiently, and (3) rounding the optimal LP solution to an integral solution.

2 Metric Facility Location Problem

Here we have a set of locations, $V = F \cup C$, where F denotes facility locations, and C denotes client locations. We are also given a metric (V, d), where d is denotes the distances between facility/client locations. Recall that we have $d_{ij} = d_{ji}$ for all $i, j \in V$ and $d_{ij} \leq d_{ik} + d_{jk}$ for all $i, j, k \in V$. For any set $S \subseteq V$ and vertex $j \in V$ we define $d(j, S) = \min_{i \in S} d_{ij}$. There is an opening cost f_i associated to each facility $i \in F$. The goal is to open a set $S \subseteq F$ facilities and connect each client j to some $i \in S$ so as to minimize the sum of opening and connection costs:

$$\min_{S \subseteq F} \quad \sum_{i \in S} f_i + \sum_{j \in C} d(j, S)$$

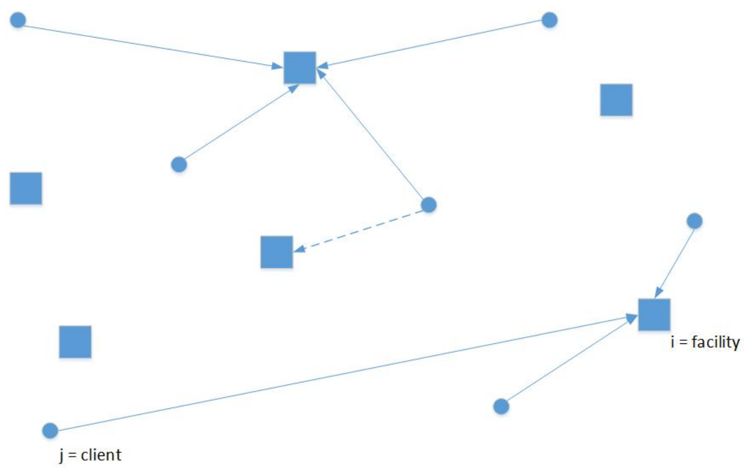

Figure 1: Example of metric facility locations

2.1 Integer Program Formulation

Now we introduce an integer programming (IP) formulation. There are two decision variables:

$$x_i = \begin{cases} 1 & \text{if facility } i \text{ is open} \\ 0 & \text{if facility } i \text{ is closed} \end{cases}$$

$$y_{ij} = \begin{cases} 1 & \text{if client } j \text{ is assigned to facility} i \\ 0 & \text{otherwise} \end{cases}$$

Then, the IP is as follows.

$$\begin{aligned} \min \quad & \sum_{i \in F} f_i x_i + \sum_{i \in C, j \in F} d_{ij} y_{ij} & \\ \text{s.t.} \quad & \sum_{i \in F} y_{ij} = 1 & \forall\ j \in C \\ & y_{ij} \le x_i & \forall\ i \in F, j \in C \\ & x_i, y_{ij} \in \{0,1\} & \forall\ i \in F, j \in C \end{aligned}$$

We obtain a linear programming relaxation from the integer program by replacing the constraints $x_i \in \{0,1\}$ and $y_{ij} \in \{0,1\}$ with $0 \le x_i \le 1$ and $0 \le y_{ij} \le 1$. Note that the optimal LP value is clearly at most the optimal value of the facility location instance. Let (x, y) denote the optimal LP solution. Then we will round (x, y) to $(\bar{x}, \bar{y})$ an integer solution. The main part of the analysis is to bound the ratio $\frac{cost(\bar{x},\bar{y})}{cost(x,y)}$.

2.2 Rounding Algorithm

For any vertex v and radius r, define the ball $B(v, r) = \{\ u \in V : d(v, u) \le r\ \}$.

Let L_j denote the the LP connection cost of client $j \in C$, we have:

$$L_j = \sum_{i \in F} d_{ij}\ y_{ij}$$
$$\sum_i y_{ij} = 1$$

We will use the ball around each client j with radius αL_j where $\alpha > 1$ will be set later. We use $B(j, \alpha L_j) \triangleq B_j$. See Figure 2.

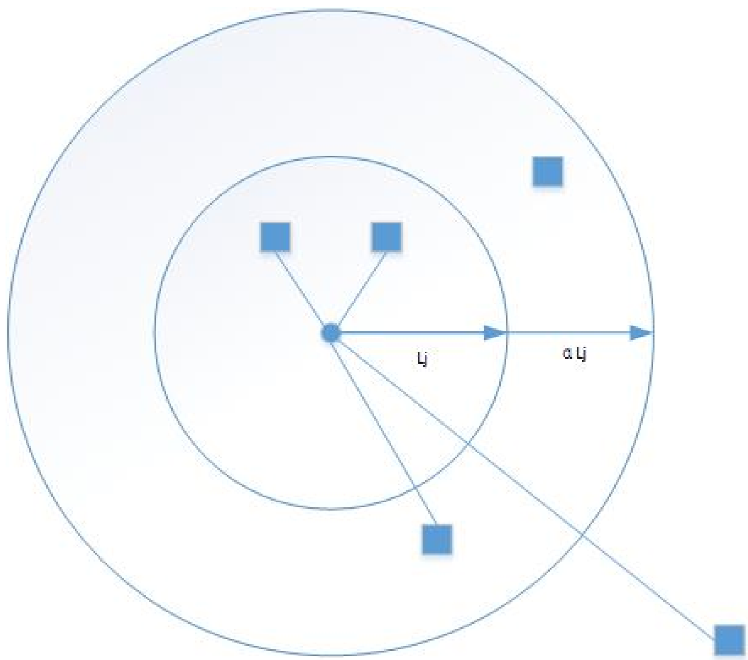

Figure 2: Relation of J_j and αL_j

Now, we can follow the procedure to find which facility to open.

(1) Sort clients as: $L_1 \leq L_2 \leq L_3 \leq ... \leq L_{|C|}$;

(2) Greedily pick a maximal set of disjoint balls in the order $1, 2, ...|C|$. Let $\{B_j, j \in I\}$ be the chosen balls;

(3) Open the cheapest facility $\pi(j)$ in each $\{B_j, j \in I\}$. Let $S = \{\pi(j) : j \in I\}$.

2.3 Analysis

Let $F^* = \sum_{i \in F} f_i x_i$ denote the facility cost in the LP solution, and $D^* = \sum_{i \in C, j \in F} d_{ij} y_{ij}$ its connection cost. We first bound the facility cost of the rounded solution.

Lemma 2.1. *The facility cost* $\sum_{i \in S} f_i \leq \frac{F^*}{1-\frac{1}{\alpha}}$.

Proof. We will show $f_{\pi(j)} \leq \frac{1}{1-\frac{1}{\alpha}} \sum_{k \in B_j} f_k y_{jk}$ for each $j \in C$. The would then follow by the disjointness of $\{B_j : j \in I\}$, by summing over all $j \in I$.

We claim that $\sum_{k\in B_j} y_{jk} \geq 1 - \frac{1}{\alpha}$ for each $j \in C$. If not, then we must have $\sum_{k\notin B_j} y_{jk} > \frac{1}{\alpha}$, which implies:

$$L_j = \sum_{i\in F} d_{ji}y_{ij} \geq \sum_{k\notin B_j} y_{kj}(\alpha L_j) > \alpha\frac{L_j}{\alpha} = L_j,$$

a contradiction!

Now we have:

$$f_{\pi(j)} = \min_{k\in B_j} f_k \leq \frac{\sum_{k\in B_j} f_k y_{jk}}{\sum_{k\in B_j} y_{jk}} \leq \sum_{k\in B_j} f_k \frac{y_{jk}}{1-\frac{1}{\alpha}}$$

As discussed above, this implies the lemma. □

Next, we bound the connection cost.

Lemma 2.2. *For each client j, the distance $d(j,S) \leq 3\alpha L_j$. So the connection cost $\leq 3\alpha D^*$.*

Proof. Suppose $j \in I$. Then $d(j,S) \leq \alpha L_j$ as we open some facility in B_j.

Now suppose $j \notin I$. Then there is some $p < j$ with $B_p \cap B_j \neq \emptyset$ and $p \in I$. Let v denote some vertex in $B_p \cap B_j$. Using triangle inequality,

$$d(j,p) \leq d(j,v) + d(v,p) \leq \alpha L_j + \alpha L_p \leq 2\alpha L_j,$$

where the last inequality used the greedy ordering on clients. This implies:

$$d(j,\pi(p)) \leq d(j,p) + d(p,\pi(p)) \leq 3\alpha L_j.$$

So in either case we have $d(j,S) \leq 3\alpha L_j$. By adding over all $j \in C$, the total connection cost is at most $3\alpha D^*$. □

Based on above lemmas, we have:

Theorem 2.1. *There is a 4-approximation algorithm for metric facility location.*

Proof. We optimize the choice of the parameter $\alpha > 1$. From the above, we know that the facility cost $\leq \frac{1}{1-\frac{1}{\alpha}}F^*$. And the connection cost $\leq 3\alpha L_j$. Then we have:

$$ALG \leq \frac{1}{1-\frac{1}{\alpha}}F^* + 3\alpha D^* \leq \max(\frac{\alpha}{\alpha-1}, 3\alpha)\cdot(F^* + D^*).$$

Setting $\alpha = \frac{1}{3} + 1$, we get $ALG \leq 4(F^* + D^*)$. We see that the overall cost is at most 4 times optimal. □