Saarland Informatics Campus

Script of Lecture 7, Approximation Algorithms
Summer term 2017

Tobias Mömke, Hang Zhou
http://www-cc.cs.uni-saarland.de/course/61/

Written by Hang Zhou

1 The Uncapacitated Facility Location Problem

Input: a set D of demands (or clients); a set F of facilities; for each client $j \in D$ and facility $i \in F$, there is a cost c_{ij} of assigning client j to facility i; and there is a cost f_i associated with each facility $i \in F$.

Output: Choose a subset of facilities $F' \subseteq F$ which minimizes the total cost of facilities in F' plus the total cost of assigning each client to the nearest facility in F', i. e., $\sum_{i\in F'} f_i + \sum_{j\in D} \min_{i\in F'} c_{ij}$. (The first term is called the *facility cost*, and the second term is called the *assignment cost.*)

Assumption: the assignment costs obey the triangle inequality. More precisely, for any clients j, l and facilities i, k, we have $c_{ij} \leq c_{il} + c_{kl} + c_{kj}$.

2 Integer Program

For each facility $i \in F$, create a variable $y_i \in \{0,1\}$: if we open the facility i, then $y_i = 1$, otherwise $y_i = 0$. For each facility $i \in F$ and for each client $j \in D$, create a variable $x_{ij} \in \{0,1\}$: if we assign client j to facility i, then $x_{ij} = 1$, otherwise $x_{ij} = 0$.

The goal:

$$\text{minimize} \sum_{i\in F} f_i y_i + \sum_{i\in F, j\in D} c_{ij} x_{ij}.$$

To ensure that each client is assigned to exactly one facility, we use the constraints:

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

To ensure that the clients are assigned to the facilities that are open, we use the constraints:

$$x_{ij} \leq y_i, \text{ for all } i \in F \text{ and } j \in D.$$

Integer program formulation:

$$\begin{aligned}
\text{minimize} & \sum_{i\in F} f_i y_i + \sum_{i\in F, j\in D} c_{ij} x_{ij} & \\
\text{subject to} & \sum_{i\in F} x_{ij} = 1 & \forall j \in D \\
& x_{ij} \leq y_i, & \forall i \in F, j \in D \\
& x_{ij} \in \{0,1\}, & \forall i \in F, j \in D \\
& y_i \in \{0,1\}, & \forall i \in F
\end{aligned}$$

3 Linear Program and its Dual

Linear program: replace the constraints $x_{ij} \in \{0,1\}$ and $y_i \in \{0,1\}$ by $x_{ij} \geq 0$ and $y_i \geq 0$.

Dual linear program: Since the dual LP should be a maximization problem, we need to design a formulation for a lower bound for the uncapacitated facility location problem.

If all the facilities have costs $f_i = 0$, then we can open all facilities. By setting $v_j := \min_{i \in F} c_{ij}$ for each $j \in D$, we have the total cost is $\sum_{j \in D} v_j$.

Now consider the general case where the costs f_i can be positive. For each facility i, we distribute the cost f_i to the clients that use the facility i: client j pays a cost of w_{ij} (in addition to its assignment cost c_{ij}) if it uses the facility i. We would like that $f_i = \sum_{j \in D} w_{ij}$ where each $w_{ij} \geq 0$. We set $v_j := \min_{i \in F}(c_{ij} + w_{ij})$. Then $\sum_{j \in D} v_j$ is a lower bound on the total cost of the uncapacitated facility location problem.

However, we do not specify a way to distribute the facility costs into $\{w_{ij}\}_{i,j}$. Since $\sum_{j \in D} v_j$ is a lower bound under any distribution of the facility costs, we regard $\{w_{ij}\}_{i,j}$ as variables in the dual LP where the goal is to maximize $\sum_{j \in D} v_j$.

Dual linear program formulation (observe the constraint-variable correspondence):

$$\begin{aligned}
\text{maximize} \sum_{j \in D} v_j & & \\
\text{subject to} \sum_{j \in D} w_{ij} = f_i, & & \forall i \in F \\
v_j \leq c_{ij} + w_{ij}, & & \forall i \in F, j \in D \\
w_{ij} \geq 0, & & \forall i \in F, j \in D
\end{aligned}$$

4 Deterministic Rounding

Let (x^*, y^*) be an optimal solution to the primal LP.

Definition 1. *We say that a client j is a* neighbor *of a facility i if $x^*_{ij} > 0$. Let $N(j) = \{i \in F : x^*_{ij} > 0\}$.*

Lemma 1. *If (x^*, y^*) is a solution to the primal LP and (v^*, w^*) a solution to the dual LP, then $x^*_{ij} > 0$ implies $c_{ij} \leq v^*_j$.*

Proof. By complementary slackness (a property of optimal linear programming solutions), $x^*_{ij} > 0$ implies $v^*_j = c_{ij} + w^*_{ij}$. The statement follows since $w^*_{ij} \geq 0$. ☐

If we can pick a set of clients $\{j_k\}_{k \geq 1}$ such that their neighborhoods $N(j_k)$ are disjoint, then we are able to pick one facility inside each $N(j_k)$ such that the total facility cost is bounded. More precisely, for each $N(j_k)$, we pick the cheapest facility i_k in $N(j_k)$. The facility cost of i_k is

$$\begin{aligned}
f_{i_k} &= f_{i_k} \sum_{i \in N(j_k)} x^*_{ij_k} & & \text{(LP constraint } \textstyle\sum_{i \in F} x_{ij} = 1\text{)} \\
&\leq \sum_{i \in N(j_k)} f_i \cdot x^*_{ij_k} & & (i_k \text{ is the cheapest facility in } N(j_k)) \\
&\leq \sum_{i \in N(j_k)} f_i \cdot y^*_i & & \text{(LP constraint } x_{ij} \leq y_i\text{)}.
\end{aligned}$$

Summing over all facilities i_k, we have the total facility cost is:

$$\sum_k f_{i_k} \leq \sum_k \sum_{i \in N(j_k)} f_i \cdot y^*_i \leq \sum_{i \in F} f_i \cdot y^*_i,$$

where the last inequality is because the neighborhoods $N(j_k)$ are disjoint subsets of F.

Our goal is to open a set of facilities with the above property such that the assignment cost is not too large.

Definition 2. *Let $N^2(j)$ denote all neighboring clients of the neighboring facilities of client j; that is $N^2(j) = \{k \in D :$ client k is a neighbor of some facility $i \in N(j)\}$.*

In Algorithm 1, we repeatedly pick a client j_k and obtain the facility i_k which has the cheapest facility cost in $N(j_k)$, we then assign j_k and all previously unassigned clients in $N^2(j_k)$ to i_k. In this way, the neighborhoods $N(j_k)$ are disjoint.

In order to bound the assignment cost, in each iteration k, we pick a client j_k which minimizes v_j^* over all unassigned clients in that iteration.

Algorithm 1 Deterministic rounding algorithm

Solve LP to get optimal primal solution (x^*, y^*) and dual solution (v^*, w^*)
$C \leftarrow D$
$k \leftarrow 0$
while $C \neq \emptyset$ **do**
    $k \leftarrow k + 1$
    Choose $j_k \in C$ that minimizes v_j^* over all $j \in C$
    Choose $i_k \in N(j_k)$ to be the cheapest facility in $N(j_k)$
    Assign j_k and all unassigned clients in $N^2(j_k)$ to i_k
    $C \leftarrow C - \{j_k\} - N^2(j_k)$

Theorem 1. *Algorithm 1 is a 4-approximation algorithm for the uncapacitated facility location problem.*

Proof. The facility cost is $\sum_k f_{i_k} \leq \sum_{i \in F} f_i \cdot y_i^* \leq \text{OPT}$. We only need to bound the assignment cost by at most 3OPT.

Consider a fixed iteration k. Let $j = j_k$ and $i = i_k$. The cost of assigning j to i is $c_{ij} \leq v_j^*$ (Lemma 1). Now consider a client $l \in N^2(j)$. Let $h \in F$ be some facility that is a common neighbor of j and l. The cost of assigning l to i is

$$\begin{aligned} c_{il} &\leq c_{ij} + c_{hj} + c_{hl} && \text{(triangle inequality on } c) \\ &\leq v_j^* + v_j^* + v_l^* && \text{(Lemma 1)} \\ &\leq 3v_l^* && \text{(by the choice of } j) \end{aligned}$$

Therefore, the total assignment cost is at most $3 \sum_{j \in D} v_j^*$, which is at most 3OPT by the weak duality. ☐

5 Randomized Rounding

In the previous rounding algorithm, we bound the facility cost by OPT and the assignment cost by 3OPT respectively. However, we observe that bound the facility cost by $\sum_{i \in F} f_i \cdot y_i^* \leq \text{OPT}$ is not tight. We want to achieve a better ratio by bounding the two terms together, using a randomized rounding algorithm.

In the randomized algorithm, once we have selected a client j, instead of opening the cheapest facility in $N(j)$, we open the facility $i \in N(j)$ with probability x_{ij}^* (note that $\sum_{i \in N(j)} x_{ij}^* = 1$).

We also modify the choice of the client j_k selected in the iteration k: First, for every $j \in D$, we define $C_j^* = \sum_{i \in F} c_{ij} \cdot x_{ij}^*$, i.e., the assignment cost of client j in the LP solution (x^*, y^*). We then choose the client j_k that minimizes $v_j^* + C_j^*$ over all unassigned clients in that iteration. See Algorithm 2.

Theorem 2. *Algorithm 2 is a randomized 3-approximation algorithm for the uncapacitated facility location problem.*

Algorithm 2 Randomized rounding algorithm

Solve LP to get optimal primal solution (x^*, y^*) and dual solution (v^*, w^*)
$C \leftarrow D$
$k \leftarrow 0$
while $C \neq \emptyset$ **do**
    $k \leftarrow k + 1$
    Choose $j_k \in C$ *that minimizes* $v_j^* + C_j^*$ *over all* $j \in C$
    Choose $i_k \in N(j_k)$ *according to the probability distribution* $x_{ij_k}^*$
    Assign j_k and all unassigned clients in $N^2(j_k)$ to i_k
    $C \leftarrow C - \{j_k\} - N^2(j_k)$

Proof. As before, the facility cost is $\sum_k f_{i_k} \leq \sum_{i \in F} f_i \cdot y_i^*$. In the following, we analyze the assignment cost.

Consider a fixed iteration k. Let $j = j_k$ and $i = i_k$. The expected cost of assigning j to i is

$$\sum_{i \in N(j)} c_{ij} \cdot x_{ij}^* = C_j^*.$$

Now consider a client $l \in N^2(j)$. Let $h \in F$ be some facility that is a common neighbor of j and l. The expected cost of assigning l to i is

$$\begin{aligned} c_{il} &\leq \left(\sum_{i \in N(j)} c_{ij} \cdot x_{ij}^* \right) + c_{hj} + c_{hl} && \text{(triangle inequality on } c\text{)} \\ &\leq C_j^* + c_{hj} + c_{hl} \\ &\leq (C_j^* + v_j^*) + v_l^* && \text{(Lemma 1)} \\ &\leq (C_l^* + v_l^*) + v_l^* && \text{(by the choice of } j\text{)} \\ &= C_l^* + 2v_l^* \end{aligned}$$

Therefore, the total expected cost of the solution is at most

$$\begin{aligned} &\sum_{i \in F} f_i \cdot y_i^* + \sum_{j \in D} (C_j^* + 2v_j^*) \\ \leq &\sum_{i \in F} f_i \cdot y_i^* + \sum_{i \in F, j \in D} c_{ij} \cdot x_{ij}^* + 2 \sum_{j \in D} v_j^* \\ \leq &3\text{OPT} \end{aligned}$$

□