COM S 783 / OR&IE 634 Approximation Algorithms   Fall Semester, 2003

Lecture 20: Nov 17, 2003

Lecturer: Yuval Rabani   *Scribe: Sumit Kunnumkal*

1 Greedy algorithm for facility location

We will look at a greedy algorithm for the uncapacitated facility location problem. It has a similar flavor to the approximation algorithm for set cover, in that it uses the method of dual fitting. The greedy approach yields some of the strongest results for the facility location problem. The algorithm presented here has an approximation factor of 1.61 and is due to Jain et al. [1]. The best known factor for this problem currently is 1.52 and is due to Mahdian, Ye and Zhang [2]. Regarding hardness results, Guha and Khuller [3] showed that the best possible approximation factor for this problem is 1.463 assuming $NP \not\subseteq DTIME[n^{O(\log\log n)}]$.

2 Algorithm

Recall that in the uncapacitated facility location problem we are given a set F of facilities and a set C of clients. There is a specified cost of opening facility f_i and a specified connection cost c_{ij} between facility $i \in F$ and client $j \in C$. We assume that the connection costs satisfy the triangle inequality. The objective is to open a subset of facilities in F to serve all the clients in C such that the total facility and connection cost is minimized.

The algorithm itself is simple to state:

1. We introduce a notion of time. The algorithm starts at time 0. We define a variable $\beta_j \ \forall j \in C$. This can be thought of as the "budget" of client j, which it uses to pay for opening facilities and getting connected to them. Let U denote the set of unconnected clients. Initially, no facility is open, all clients are unconnected ($U = C$) and $\beta_j = 0 \ \forall j \in C$.

   At every moment, each client j bids for every unopened facility i. Client j's bid is calculated as follows: If j is unconnected, then its bid to facility i is $\max\{\beta_j - c_{ij}, 0\}$. So an unconnected client j bids the the remainder of its current budget (left after paying for the connection cost to facility i), towards opening a facility. On the other hand, if j is already connected to some facility say i', then its bid to facility i will be $\max\{c_{i'j} - c_{ij}, 0\}$, i.e., the amount it would save by switching to facility i.

2. While $U \neq \emptyset$, increase the time, and uniformly raise the budget β_j for each $j \in U$, until one of the following events occur:

(a) For some unopened facility i, the total bid on it is equal to its opening cost f_i. In this case, open facility i and connect to it all clients that made a positive bid towards it. Update the set of unconnected clients U.

(b) For some unconnected client j and some open facility i, $\beta_j = c_{ij}$. In this case, connect j to i and remove j from U.

Remark 1: Observe that once a client gets connected to some facility, its budget is frozen. Let α_j denote this final value.

Remark 2: Once client j makes a positive bid towards opening some facility, it cannot take back its bid at a later time. All it can bid towards another facility is the savings in connection cost.

Remark 3: Although the β_j's have been described being raised in a continuous manner, the process can clearly be discretized.

3 Analysis

The analysis relies on the fact that a feasible dual solution provides a lower bound on the optimal cost of the primal LP (when the primal is a minimization problem and the dual is a maximization problem), which in turn is a lower bound on the optimal integer solution. The idea is to show that the solution generated by the algorithm when scaled by a suitable factor γ gives a feasible dual solution. In that case, we clearly have a γ approximation. We get a bound on γ by solving a series of factor-revealing LPs. We start with the following fact:

Claim 1. The total cost of the solution found by the algorithm $= \sum_{j \in C} \alpha_j$, where α_j is the final value of client j's budget.

This can be easily seen since each client pays for its connection cost and each opened facility is fully paid for as clients cannot take back their bids from an opened facility (Remark 2).

Now, we consider an LP relaxation of the facility location problem and its dual. We say that a *star* consists of one facility and several clients. The cost of a star S, $c(S)$, is the sum of the facility opening cost and the connection costs of all clients in the star, i.e., $c(S) = f_i + \sum_{j \in S} c_{ij}$. Let $\mathcal{S}$ denote the set of all stars. The facility location problem can then be thought of as picking a minimum cost subset of stars such that each client is at least in one star. So, we get the following IP for the facility location problem, which has an indicator variable x_S for each star $S \in \mathcal{S}$.

$$\begin{array}{lll} \min & \sum_S c_S x_S & \\ \text{s.t} & \sum_{S:\ j \in S} x_S \geq 1 & \forall j \in C \\ & x_S \in \{0,1\} & \forall S \in \mathcal{S} \end{array}$$

The LP relaxation is obtained by replacing $x_S \in \{0,1\}$ by $x_S \geq 0$. Its dual is given by:

$$\begin{array}{rll} \max & \sum_{j \in C} \alpha_j & \\ \text{s.t} & \sum_{j \in S} \alpha_j \leq c_S & \forall S \in \mathcal{S} \\ & \alpha_j \geq 0 & \forall j \in C \end{array}$$

Now, we want to find a number γ so that the solution of the greedy algorithm scaled by γ is dual feasible, i.e., for every star S, $\sum_{j \in S} \alpha_j \leq \gamma c_S$. We express various constraints that are imposed by the algorithm or by the structure of the problem as inequalities and get a bound on γ by solving a series of linear programs (factor-revealing LPs).

Consider a star S consisting of facility i and k clients numbered 1 through k. Wlog we can assume that the clients are sorted so that $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_k$. Consider the moment t, just before client j gets connected for the first time. At this time, clients $1 \ldots j-1$ might be connected to some facility. So, for all clients $j' < j$ which are connected to some facility at this time, let $r_{j'j}$ denote its connection cost. Otherwise, if $j' < j$ is unconnected, let $r_{j'j} = \alpha_{j'}$. Observe that the second case occurs iff $\alpha_{j'} = \alpha_j$. At this time, the amount each client bids on facility i is:

- $\max\{r_{j'j} - c_{ij'}, 0\}$ for $j' < j$
- $\max\{t - c_{ij'}, 0\}$ for $j' \geq j$

Note that clients $j' \geq j$ are unconnected at time t and so bid the difference between their current budget, $t = \alpha_j$, and the connection cost to facility i, $c_{ij'}$. Now, since the total bid to a facility at any time cannot exceed its opening cost, we get the following set of constraints $\forall 1 \leq j \leq k$:

$$\sum_{j'=1}^{j-1} \max\{r_{j'j} - c_{ij'}, 0\} + \sum_{j'=j}^{k} \max\{\alpha_j - c_{ij'}, 0\} \leq f_i \tag{1}$$

We will obtain the next set of constraints using the triangle inequality. Consider some client $j' < j$. If j' is connected to some facility, say i' at time t (when j gets connected for the first time), then it must be that $\alpha_j \leq c_{i'j}$. If this were not the case, then j would have connected to the open facility i' at an earlier time, which is a contradiction. So, for every $1 \leq j' < j \leq k$,

$$\alpha_j \leq c_{i'j} \leq r_{j'j} + c_{ij'} + c_{ij} \tag{2}$$

where, the second inequality comes from the metric assumption and by noting that $r_{j'j}$ is connection cost of j' to facility i' (by definition). Also, note that if j' is unconnected at time t, then $\alpha_j = \alpha_{j'} = r_{j'j}$ and so the inequality holds trivially.

Finally, note that once a client gets connected to some facility, it can only switch to another facility if the connection cost is lower. This implies that for every j:

$$r_{j,j+1} \geq r_{j,j+2} \geq \ldots \geq r_{j,k} \tag{3}$$

The above set of inequalities (1, 2, 3) form the following factor-revealing LP:

$$\begin{array}{lll}
\text{maximize} & \frac{\sum_{j=1}^{k} \alpha_j}{f+\sum_{j=1}^{k} d_j} & \\
\text{s.t} & \alpha_j \leq \alpha_{j+1} & \forall 1 \leq j < k \\
 & r_{j'j} \geq r_{j'j+1} & \forall 1 \leq j' < j < k \\
 & \alpha_j \leq r_{j'j} + d_{j'} + d_j & \forall 1 \leq j' < j \leq k \\
 & \sum_{j'=1}^{j-1} \max\{r_{j'j} - d_{j'}, 0\} + \sum_{j'=j}^{k} \max\{\alpha_j - d_{j'}, 0\} \leq f & \forall 1 \leq j \leq k \\
 & r_{j'j}, \alpha_j, d_j, f \geq 0 &
\end{array}$$

Notice that:

- By introducing new variables and inequalities, the non-linearities can be eliminated.
- The connection cost of clients and the facility opening cost are represented by non-negative variables — d_j, f, respectively.
- Every star S with a facility and k clients in the solution of the algorithm satisfies the constraints in the above LP and so is a feasible solution.

Thus, we have the following lemma:

Lemma 2. If z_k denotes the solution of the factor-revealing LP, then for every star S in the algorithm consisting of a facility and k clients, $\sum_{j \in S} \alpha_j \leq z_k c_S$.

Claim 1 and Lemma 2 implies the following:

Lemma 3. Let z_k be the solution of the factor-revealing LP, and $\gamma := \sup_k\{z_k\}$. Then the greedy algorithm solves the metric facility location problem with an approximation factor of γ.

Jain et al. [1] show that $\gamma \leq 1.61$. We will not go into the details of the proof.

References

[1] K. Jain, M. Mahdian, E. Markakis, A. Saberi and Vijay V. Vazirani. Greedy Facility Location Algorithms Analyzed using Dual Fitting with Factor-Revealing LP. *Journal of the ACM*, 50(6), pp. 795-824, November 2003.

[2] M. Mahdian, Y. Ye and J. Zhang. Imrpoved approximation algorithms for metric facility location problems. In *Proceedings of 5th International Workshop on Approximation Algorithms for Combinatorial Optimization (APPROX 2002)*, 2002.

[3] S. Guha and S. Khuller. Greedy strikes back: Improved facility location algorithms. *Journal of Algorithms*, 31:228-248, 1999.

[4] K. Jain, M. Mahdian and A. Saberi. A new greedy approach for facility location problems. In *ACM Symposium on Theory of Computing*, 2002.