COMPSCI 632: Approximation Algorithms	September 27, 2017
Lecture 8	
Lecturer: Debmalya Panigrahi	*Scribe: Alex Steiger*

1 Overview

Recall the ***uncapacitated facility location problem (UFLP)***: given a set of facilities F, clients C, opening costs f_i for each $i \in F$, and a metric $d(\cdot,\cdot)$ over F and C, find a subset of facilities $F' \subseteq F$ such that the sum of their opening costs and the sum of distances from each client to its closest facility in F' is minimized. That is, find $\arg\min_{F' \subseteq F} \left(\sum_{i \in F'} f_i + \sum_{j \in C} \min_{i \in F'} d_{ij}\right)$.

In this lecture, we will use the dual fitting method to design an approximation algorithm for the UFLP. We previously gave a 3-approximation algorithm for this problem using the primal-dual method in a previous lecture. Dual fitting is similar to the primal-dual method in that solutions to the primal and dual LPs are constructed without solving either of them directly, but differs in that both solutions constructed will have the same cost and only the primal is guaranteed to be feasible. The dual solution is then "fit" to satisfy the dual constraints by scaling its variables, then its scaled cost is used as a surrogate between the cost of the primal and optimal solutions.

1.1 Notation

For convenience, let $(x)^+ = \max\{x, 0\}$ for any quantity x. Additional notation particular to each algorithm will be defined in their respective sections.

2 A 2-approximation for the UFLP

We will actually present three algorithms. In each of them we have three common variables: a subset of opened facilities $F' \subseteq F$ which is initially set to $\varnothing$, an explicit assignment of clients to opened facilities $g : C \to F'$ which is initially undefined for all clients, and a subset of clients not yet assigned to any opened facility $X \subseteq C$ which is initially set to C.

2.1 A starting point

The first algorithm closely resembles the greedy $O(\log n)$-approximation algorithm for set cover which repeatedly picks the set that minimizes the ratio of its cost to the number of its uncovered elements until all elements are covered. The idea is to extend this approach and greedily open an unopened facility $i \in F - F'$ and assign a subset of unconnected clients $Y \subseteq X$ to it which minimize the ratio of the opening and connection costs incurred and the number of clients in Y. That is, while X is non-empty (and thus there are unconnnected clients), choose

$$i, Y = \arg\min_{\substack{i \in F - F' \\ Y \subseteq X}} \frac{f_i + \sum\limits_{j \in Y} d_{ij}}{|Y|},$$

then open i and assign each client in Y to it (which is reflected by updating F', X, and g accordingly).

To see that this algorithm can be implemented to run in polynomial time although there are exponentially many subsets $Y \subseteq X$, we note that it is sufficient to consider, for each choice of $i \in F - F'$, only the subsets composed of the ℓ clients closest to Y for each $\ell \in \{1, \ldots, |X|\}$. Thus, we only have to consider a polynomial number of subsets in each iteration which can easily be obtained in polynomial time, so the algorithm can be implemented to run in polynomial time overall.

Next, instead of analyzing this algorithm, we note two simple improvements from which we will obtain the second algorithm that we describe in detail in the next section. The first is to allow facilities to be chosen again after it is opened so that clients can be assigned to it in later iterations but without incurring the opening cost again. To implement this, we will set the facility cost of opened facilities to zero and consider every $i \in F$ instead of those in $F - F'$ in each iteration. The second improvement is to allow clients to be reassigned in future iterations, and to consider the resulting potential savings when greedily choosing which facility i to open (if it is not already due to the previous improvement). Consider a facility i and a client j connected to it for the first time at some point in the algorithm. It is possible that in a later iteration a different facility i' is opened that is closer to j than i, i.e. $d_{ij} > d_{i'j}$, so reassigning j to i' would decrease its connection cost. It follows that when choosing a facility i in each iteration, we should consider the potential savings in connection costs $(d_{g(j)j} - d_{ij})^+$ for each $j \notin X$ if i were to be opened, and if it is chosen, we should greedily reassign those whose difference in those connection costs are positive. Recall that $g(j)$ is the facility to where client $j \notin X$ is currently assigned in that iteration.

2.2 An improved second algorithm

The second algorithm reflects the improvements to the first algorithm from the previous section and is as follows. We additionally add the 0'th step which is significant only in the analysis and will be discussed in detail later.

0. We artificially scale the opening costs of each facility f_i by a factor of $\gamma \geq 1$ to be determined later. Denote the scaled costs by $\widehat{f}_i = \gamma f_i$. As mentioned, this step will be significant only in the analysis.

1. While X is not empty, choose the facility $i \in F$ and subset $Y \subseteq X$ that minimizes

$$\frac{\widehat{f}_i - \sum_{j \notin X} (d_{g(j)j} - d_{ij})^+ + \sum_{j \in Y} d_{ij}}{|Y|}.$$

   Then open i if it is not opened already, assign each $j \in Y \cup \left\{ j \notin X \;\middle|\; d_{g(j)j} > d_{ij} \right\}$ to i, and set $\widehat{f}_i$ to zero (updating F', g, and X accordingly).

For the same reason as in the first algorithm, although there are an exponential number of subsets $Y \subseteq X$, it is sufficient to consider a polynomial number of them, and thus this algorithm can be implemented in polynomial time.

To analyze this algorithm, we assert its equivalence to the following third algorithm which constructs an explicit solution to the dual of the LP relaxation for this problem while constructing F' and g, then analyze that instead. Proving the equivalence of these algorithms is left as an exercise.

2.3 A third equivalent algorithm

First, let us recall the dual LP for the UFLP:

$$\begin{aligned}
\max \quad & \sum_{j \in C} \alpha_j \\
\text{s.t.} \quad & \sum_{j \in C} \beta_{ij} \leq f_i && \forall i \in F \\
& \alpha_j - \beta_{ij} \leq d_{ij} && \forall i \in F, j \in C \\
& \alpha_j, \beta_{ij} \geq 0 && \forall i \in F, j \in C
\end{aligned}$$

For convenience, we rewrite it (with non-linear constraints) by implicitly setting $\beta_{ij} = (\alpha_j - d_{ij})^+$ for all $i \in F, j \in C$:

$$\begin{aligned}
\max \quad & \sum_{j \in C} \alpha_j \\
\text{s.t.} \quad & \sum_{j \in C} (\alpha_j - d_{ij})^+ \leq f_i && \forall i \in F \\
& \alpha_j \geq 0 && \forall j \in C
\end{aligned}$$

The algorithm is as follows:

0. As in the first algorithm, we artificially scale the opening costs of each facility f_i by a factor of $\gamma \geq 1$ to be determined later and denote the scaled costs by $\widehat{f_i} = \gamma f_i$.

1. In addition to F', g, and X, we additionally define α_j for all $j \in C$, all initialized to zero, which are the variables of the dual LP. We define the ***bid*** of a client on a facility i to be $(\alpha_j - d_{ij})^+$ if $j \in X$, and $(d_{g(j)j} - d_{ij})^+$ if $j \notin X$ and so it is already connected to some facility $g(j)$.

2. Then we uniformly increase the α_j's of all clients $j \in X$ until one of the following events occurs, or X is empty. If more than one event occurs simultaneously, they are handled in an arbitrary order.

   (a) If the sum of ***positive*** bids on a facility $i \notin F'$ equals its opening cost $\widehat{f_i}$, we open i and connect all clients with the positive bids on it (and update F', g, and X accordingly). The clients in X with non-negative bids have their α_j's frozen, i.e. they will no longer be increased throughout the remainder of the algorithm.

   (b) If $\alpha_j = d_{ij}$ for some unconnected client $j \in X$ and opened facility $i \in F'$, we assign j to i, remove i from X, and freeze α_j.

The main idea is that when a client j is connected for the first time at time α to some facility i, they pay for their own connection cost d_{ij}, and in the first type of event, they also pay for a portion of i's opening cost $\widehat{f_i}$ with their positive bid $\alpha - d_{ij}$. The remainder of $\widehat{f_i}$ is paid for by the positive bids of already connected clients, which is effectively the redistribution of cost paid from their original connection cost. Note that the bids of already connected clients are equal to exactly the savings incurred if i was opened. Furthermore, like the second algorithm, when a facility i is first opened, each client $j \notin X$ closer to i than where they are currently connected $g(j)$ is reconnected to i.

Fact 1. *Immediately after each event is handled, the sum of frozen α_j's is equal to the opening costs of the facilities in F' and the connection costs $d_{g(j)j}$ for each $j \notin X$.*

Proof. This is directly implied by the fact that the cost incurred after each event is exactly equal to the sum of (equal) α_j's of the clients newly connected in that event which we prove next. Consider any time α at

which an event occurred and let Δ be the cost incurred after the event. If the event is of the second type, then $\alpha = d_{ij}$ for an unconnected client j and opened facility i, so $\Delta = d_{ij} = \alpha$.

Otherwise, the event is of the first type. Let i be the facility opened in the event, and $A = \left\{ j \in X \;\middle|\; \alpha_j \geq d_{ij} \right\}$ and $B = \left\{ j \notin X \;\middle|\; d_{g(j)j} \geq d_{ij} \right\}$ be the unconnected and already connected clients with positive bids on i at time α, respectively. The cost incurred after the event is $\Delta = \widehat{f}_i - \sum_{j \in B}(d_{g(j)j} - d_{ij}) + \sum_{j \in A} d_{ij}$, and the event occurred because $\widehat{f}_i = \sum_{j \in B}(d_{g(j)j} - d_{ij}) + \sum_{j \in A}(\alpha_j - d_{ij})$. Substituting the latter into the former, we have $\Delta = \sum_{j \in A}(\alpha_j - d_{ij}) + \sum_{j \in A} d_{ij} = \sum_{j \in A} \alpha_j = \alpha |A|$.

□

To finish the analysis of this algorithm, we show that for a particular choice of γ, the cost of the dual solution α is within twice the optimal LP value. The crux of the proof relies on the following lemma.

Lemma 2. *For any $S \subseteq C$ and $i \in F$, $\sum\limits_{j \in S} (\alpha_j - 2d_{ij}) \leq \widehat{f}_i$.*

Proof. Let $p = |S|$ and the clients in S be ordered in non-decreasing order by their α_j's, i.e. $\alpha_1 \leq \alpha_2, \ldots, \leq \alpha_p$. Fix some client $j \in \{1, \ldots, p\}$, then consider any client $k \in \{1, \ldots, p\}$. If $k < j$, then there are two cases: k may have been connected to some previously opened facility $g(k)$ at time α_j, or it was first connected at time α_j, i.e. $\alpha_k = \alpha_j$. We define r_{jk} to be $d_{g(k)k}$ in the former case and $\alpha_k = \alpha_j$ in the latter case. Then the bid by k on i at time α_j is equal to $(r_{jk} - d_{ik})^+$. If $k \geq j$, then k is not connected to any facility before time α_j, so its bid on i at that time is $(\alpha_j - d_{ik})^+$. Thus, we have:

$$\sum_{k<j} (r_{jk} - d_{ik})^+ + \sum_{k \geq j} (\alpha_j - d_{ij})^+ \leq \widehat{f}_i \tag{1}$$

since the algorithm ensures that the total bids on any facility do not exceed its cost, and thus the bids of no subset of clients S may exceed its cost.

Now reconsider a client $k < j$. If $\alpha_k < \alpha_j$, then $r_{jk} = d_{g(k)k}$ for some facility $g(k)$ opened before time α_j. It must be the case that $\alpha_j \leq d_{g(k)j}$, otherwise client j would have a non-negative bid on $g(k)$ before α_j and would have been connected earlier. Here we apply the triangle inequality for the first and last time: $d_{g(k)j)} \leq d_{ij} + d_{ik} + d_{g(k)k}$. Then $r_{jk} = d_{g(k)k} \geq \alpha_j - d_{ij} - d_{ik}$. On the other hand, if $\alpha_k = \alpha_j$, then $r_{jk} = \alpha_k = \alpha_j$, so clearly $r_{jk} \geq \alpha_j - d_{ij} - d_{ik}$. Thus, this inequality holds for all $k < j$.

This allows us to lower bound the sums of bids by first dropping the max functions on each term, then applying the above lower bounds on r_{jk} for each $k < j$. That is, we have:

$$\sum_{k<j} (r_{kj} - d_{ik})^+ + \sum_{k \geq j} (\alpha_j - d_{ij})^+ \leq \widehat{f}_i \tag{2}$$

$$\implies \sum_{k<j} (r_{kj} - d_{ik}) + \sum_{k \geq j} (\alpha_j - d_{ij}) \leq \widehat{f}_i \tag{3}$$

$$\implies \sum_{k<j} (\alpha_j - d_{ij} - 2d_{ik}) + \sum_{k \geq j} (\alpha_j - d_{ij}) \leq \widehat{f}_i \tag{4}$$

$$\tag{5}$$

To obtain the claim set out, we sum all of these inequalities for each value of $j \in \{1, \ldots, p\}$:

$$\sum_{j=1}^{p}\left(\sum_{k<j}(\alpha_j - d_{ij} - 2d_{ik}) + \sum_{k\geq j}(\alpha_j - d_{ij})\right) \leq p\cdot\widehat{f}_i \tag{6}$$

$$\implies \sum_{j=1}^{p}(p\alpha_j - (2p-1)d_{ij}) \leq p\cdot\widehat{f}_i \tag{7}$$

$$\implies \sum_{j=1}^{p}(\alpha_j - 2d_{ij}) \leq \widehat{f}_i \tag{8}$$

The second inequality follows from the following observations. Clearly there are exactly p α_j's for each j. Fix some $u \in \{1,\ldots,p\}$. First, observe that the second term of the first nested sum contributes exactly $u-1$ d_{iu}'s when $j = u$. Then observe for each $j \in \{u+1,\ldots,p\}$, the third term of the first nested sum contributes two d_{iu}'s, and for each $j \in \{1,\ldots,u\}$, the second term of the second nested sum contributes one d_{iu}. In total, there are $(u-1)+2(p-u)+u = 2p-1$ d_{iu}'s. □

With this lemma in hand, we finish the analysis. If we divide both sides of the inequality in the lemma by 2, we have $\sum_{j\in S}\alpha_j/2 - d_{ij} \leq \widehat{f}_i/2 = \gamma f_i/2$ for any $S \subseteq C$ and $i \in F$. Now consider the dual solution α' where $\alpha'_j = \alpha_j/2$ for each $j \in C$. If we set γ to 2, we have $\sum_{j\in S}\alpha'_j - d_{ij} \leq f_i$, which we use to imply α' is dual feasible. Consider $D = \left\{ j \in C \;\middle|\; \alpha'_j - d_{ij} > 0 \right\}$. Since the previous inequality holds for *every* subset of clients, we have

$$\sum_{j\in C}(\alpha'_j - d_{ij})^+ = \sum_{j\in D}\alpha'_j - d_{ij} \leq f_i \tag{9}$$

and thus each dual constraint of the LP is satisfied for the *original* problem instance without the scaled facility costs. Let OPT be the optimal LP value and ALG be the cost of the solution F' output by the algorithm. Since α' is feasible, we have $\sum_{j\in C}\alpha'_j \leq OPT$. Thus we have

$$ALG \leq \sum_{f\in F'}\widehat{f}_i + \sum_{j\in C} d_{g(j)j} = 2\sum_{f\in F'} f_i + \sum_{j\in C} d_{g(j)j} = \sum_{j\in C}\alpha_j = 2\sum_{j\in C}\alpha'_j \leq 2\cdot OPT \tag{10}$$

where g is the assignment of clients when the algorithm terminates, and so we conclude the algorithm is a 2-approximation by setting γ to 2.

2.4 Further results

For completion, we note that the first algorithm presented actually has an approximation ratio at most 1.861, and the second (and third) have an approximation ratio at most 1.61; both of these results are shown by Jain *et al.* in [JMM+03]. In 2011, Li [Li11] gave the current best approximation algorithm with a ratio of 1.488 for this problem. The best hardness result is by Guha and Khuller [GK99] who showed it cannot be approximated better within a factor of 1.463.

3 4-approximation for k-median

In a previous lecture we showed that a 2α-approximation algorithm for the k-median problem can be obtained from any ***Lagrangian approximation preserving*** (LMP) α-approximation algorithm for UFPL. We

recall that an α-approximation algorithm A for UFLP is LMP if $\alpha \sum_{i \in F'} f_i + \sum_{j \in C} \min_{i \in F} d_{ij} \leq \alpha \cdot OPT$ where F' is the subset of facilities opened by A. Indeed, inequality 10 at the end of the previous section implies the algorithm is LMP.

4 Summary

In this lecture, we gave a 2-approximation algorithm for the UFLP based on the dual fitting technique. Since it is Lagrangian approximation preserving, this immediately implies a 4-approximation algorithm for the k-Median Problem.

References

[GK99] Sudipto Guha and Samir Khuller. Greedy strikes back: Improved facility location algorithms. *Journal of algorithms*, 31(1):228–248, 1999.

[JMM+03] Kamal Jain, Mohammad Mahdian, Evangelos Markakis, Amin Saberi, and Vijay V Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing lp. *Journal of the ACM (JACM)*, 50(6):795–824, 2003.

[Li11] Shi Li. A 1.488 approximation algorithm for the uncapacitated facility location problem. *Automata, languages and programming*, pages 77–88, 2011.