CSC2421h - Approximation Algorithms
Lecture #11

Lecturer: Allan Borodin
Scribe: Matei David

April 6, 2005

Today, we will focus on the Uncapacitated Facility Location (UFL) problem, specifically on the case where the distance function is a metric. We will see how the UFL problem is related to the k-median problem and to the Set Cover problem.

1 The Uncapacitated Facility Location problem

In the Uncapacitated Facility Location problem, we are given a set of facilities (e.g. hospitals) F, each with an opening cost f_i, and a set of cities (or clients) C, together with a distance function c_{ij} specifying the cost (or distance) of connecting each city $j \in C$ to each facility $i \in F$. The goal is to connect every city to an open facility in such a way as to minimize the total cost. This total cost consists of an opening cost and a connection cost.

Once we decide which facilities to open, we always connect a city to the nearest facility. Hence, the cost of a solution is completely determined by the set of open facilities.

We will focus on the *disjoint model*, where facilities and cities are disjoint, i.e. $F \cap C = \emptyset$. In this case, the connection graph is bipartite. An alternative is the *complete model*, where facilities and cities are the same, i.e. $F = C$. Note that any algorithm for the disjoint model also work in the complete model.

The distance function is a *metric* if it satisfies $c_{ik} \leq c_{ij} + c_{i'j} + c_{i'k}$ for all cities j, k and facilities i, i'. In general, the distance function can be arbitrary.

In the *capacitated* case, there are bounds on the number of cities allowed to connect to each facility. Note that in the capacitated case, the set of open facilities does not completely determine the solution. This case is provably harder, in the sense that the lower bounds for CFL are greater than the ratios already achieved for UFL. We concentrate on the uncapacitated case.

IP/LP Formulation We formulate UFL as an integer program by introducing 0/1 variables $x_{ij} = 1$ iff city $j \in C$ is connected to facility $i \in F$; and y_i which is 1 iff facility $i \in F$ is open. Constraints of the form $\sum_{i \in F} x_{ij} \geq 1$ specify that city j is connected to some facility. Constraints of the form $y_i - x_{ij} \geq 0$ say that city j cannot be connected to facility i unless i is open. To sum it up, the IP is

$$\min \sum_{i \in F, j \in C} c_{ij} x_{ij} + \sum_{i \in F} f_i y_i$$

subject to

$$\sum_{i \in F} x_{ij} \geq 1, \text{ for all } j \in C$$

$$y_i - x_{ij} \geq 0, \text{ for all } i \in F, j \in C$$

In the IP, we require that $x_{ij}, y_i \in \{0, 1\}$. The LP relaxation only says that $x \geq 0, y \geq 0$.

A bit of history For a while, the approximation ratio of this problem was not a constant. In the early 80's, the connection to the Set Cover problem was discovered. The Set Cover has a $\log n$ approximation, and we can also get a $\log n$ approximation for general Facility Location. In the mid 90's, Shmoys, Tardos and Aardal give the first constant approximation to the metric case, using IP/LP relaxation and rounding.

Later, Jain and Vazirani gave a primal-dual algorithm which achieves a 3-approximation. Moreover, this algorithm can be used to get a 6 (and later 4) approximation for the related k-median problem. Neither UFL nor k-median are covering problems and thus their duals are not packing problems. Historically, this is one of the first instances where a primal-dual algorithm was developed for a non-covering problem.

Following Jain and Vazirani, a sequence of Greedy (priority) algorithms achieved constant ratios 1.86, 1.61 and the currently best known ratio is 1.52. These are analyzed by the method of dual fitting. They are cleaner to state, somewhat more involved to analyze, and they dominate the previous algorithms. However, the Jain and Vazirani algorithm introduces some nice ideas and we will start with this algorithm.

The primal-dual algorithm We have previously seen two primal-dual algorithms for covering problems, where we kept primal slackness conditions tight and we relaxed dual conditions. In this algorithm, we keep dual slackness tight, and relax the primal. Moreover, the cleanup phase is a bit different. We have suggested earlier that stack algorithms model primal-dual algorithms. This is a debatable statement, as it is not clear how such algorithms can be adapted so as to model this algorithm.

Consider the LP relaxation, and think of the dual problem as a lower bound on the value of the optimal solution. We introduce coefficients α_j and β_{ij}, by which we multiply our original constraints:

$$\alpha_j \cdot \left(\sum_{i \in F} x_{ij} = 1 \right), \text{ for all } j \in C$$

$$\beta_{ij} \cdot (y_i - x_{ij} \geq 0), \text{ for all } i \in F, j \in C.$$

Note that equality constraints can be replaced by the inequality constraints ≥ 1, so that α is also non-negative. We get the following dual problem:

$$\max \sum_{j \in C} \alpha_j$$

$$\text{subject to}$$

$$\alpha_j - \beta_{ij} \leq c_{ij}, \text{ for all } i \in F, j \in C$$

$$\sum_{j \in C} \beta_{ij} \leq f_i, \text{ for all } i \in F$$

$$\alpha \geq 0, \beta \geq 0$$

As discussed in the Vazirani text, we can think of the quantity α_j as the total amount paid by city j for a solution. This consists of a connection cost, and a contribution to opening a facility. In some dual feasible solution, it might be that a city is charged for more than one facility. During the clean-up phase, we eliminate this situation.

The algorithm consists of two phases. In *phase 1*, we create a dual feasible solution, while having all dual slackness conditions tight. At the same time, we create a partial (tentative) primal solution, which might not be (integrally) feasible yet. During *phase 2*, we close down some facilities opened in phase 1, and create a primal feasible solution. For this primal solution, relaxed slackness conditions will hold with a factor of 1/3.

Slackness conditions For optimal primal and dual solutions, the following hold.

$$x_{ij} > 0 \Rightarrow \alpha_j - \beta_{ij} = c_{ij}$$

$$y_i > 0 \Rightarrow \sum_{j \in C} \beta_{ij} = f_i$$

$$\alpha_j > 0 \Rightarrow \sum_{i \in F} x_{ij} = 1$$

$$\beta_{ij} > 0 \Rightarrow y_i = x_{ij}$$

This first two are called primal slackness conditions, and the latter two are dual slackness conditions. It suffices to show that the following relaxed primal slackness conditions hold, in order to get an approximation algorithm.

$$x_{ij} > 0 \Rightarrow \frac{1}{3} c_{ij} \leq \alpha_j - \beta_{ij} \leq c_{ij}$$

$$y_i > 0 \Rightarrow \frac{1}{3} f_i \leq \sum_{j \in C} \beta_{ij} \leq f_i$$

Note that $\alpha_j - \beta_{ij} \leq c_{ij}$ and $\sum_{j \in C} \beta_{ij} \leq f_i$ simply because α, β is a dual feasible solution. In fact, the algorithm gets tight primal slackness conditions on y_i variables, and relaxed only on x_{ij} variables.

When $x_{ij} > 0$, we can think of city j being connected to facility i. There are two types of connections made by the algorithm. There are *direct connections*, where the primal slackness condition will be tight, $\alpha_j - \beta_{ij} = c_{ij}$; and there are *indirect connections*, for which we get $c_{ij} \leq 3\alpha_j$. This will suffice to show $\sum c_{ij} x_{ij} + 3 \sum f_i y_i \leq 3 \sum \alpha_j$. Since α is dual feasible, $\sum \alpha_j \leq OPT_f$, and therefore $\sum c_{ij} x_{ij} + 3 \sum f_i y_i \leq 3 \cdot OPT_f \leq 3 \cdot OPT$. Note that it's enough to have $\sum f_i y_i$ without the extra factor of 3 for UFL. However, we need the stronger bound for the related k-median problem.

Phase 1 The algorithm in phase 1 constructs a dual solution and the start of a primal solution. We are not committed to any connections made in this phase.

```
while there exist cities which are not connected
   for each city j not connected
      raise all such α_j simultaneously            % as in Steiner forest algorithm
      until α_j = c_ij for some j's and some i's         % now (i,j) is a tight edge
      if i already opened then
         connect j to i
      else
         start raising β_ij at same rate            % so that α_j − β_ij = c_ij
         until for some i (or many) Σ_j β_ij = f_i
         open facility i
         for any j such that (i,j) is tight, connect j to i          % we will not raise α_j any more
```

Remarks about phase 1.

- Eventually, every city gets connected.
- We are not worrying here about running time, which is clearly polynomial.
- If city j becomes tight at some time with more than two open facilities, resolve the connection arbitrarily.
- Dual feasibility is always maintained.

Phase 2 We now need to build a primal solution x_{ij}, y_i. Note that a certain city may be connected to two or more facilities. Consider the bipartite graph G of tight edges connecting cities and open facilities. We create another graph $H = (V, E)$, where $V \subseteq F$ is the set of open facilities, and E consists of edges (i', i) such that there exists a city j for which (i', j) and (i, j) are both tight *special* edges. An edge is called special if it is tight and $\beta_{ij} > 0$. There is an edge between open facilities i', i in H iff i', i are at distance 2 in G. For the permanent set of open facilities, take the lexicographically first maximal independent set in H (in terms of time opened).

Doing so may close some facilities, and some cities might be left unconnected. For example, suppose (i', j') and (i, j') are tight, and i' is opened before i. Then we close i. Any city j which previously depended on the connection to i, will now have to reconnect to i'. Here, we can use the triangle inequality, $c_{i'j} \le c_{i'j'} + c_{ij'} + c_{ij}$, which introduces the factor of 3 in the calculation.

The above discussion is a bit informal. A more formal presentation can be found in the book. In there, the cost paid by city j consists of two parts, $\alpha_j = \alpha_j^F + \alpha_j^C$, where α_j^F is the cost paid to open a facility, and α_j^C is the connection cost to some facility. If city j' is directly connected to facility i', $\alpha_{j'}^F = \beta_{i'j'}$ and $\alpha_j^C = c_{i'j'}$. If city j is indirectly connected to facility i', $\alpha_j^F = 0$ and $\alpha_j^C = \alpha_j$. By the properties of the metric, when j switches connections from i to i', $c_{i'j} \le c_{i'j'} + c_{ij'} + c_{ij}$. It is possible to argue that every one of the three terms on the right is at most α_j, by using the fact that facility i was opened before i'.

2 The k-Median problem

The k-median problem resembles an instance of UFL where $F = C$. Informally, we are given a set of cities, and we have to pick at most k of them, in such a way that we minimize the distance of any unselected city to a selected one. Clearly, if there are at most k cities, the optimal solution is to pick them all. If there are more than k cities, we will never want to pick less than k of them.

The IP formulation

$$\min \sum_{i,j \in C} c_{ij} x_{ij}$$

$$\text{subject to}$$

$$\sum_i y_i \le k \text{ or, equivalently } -\sum_i y_i \ge -k$$

$$\sum_i x_{ij} = 1, \text{ for all } j \in C$$

$$y_i - x_{ij} \ge 0, \text{ for all } i, j \in C$$

As before, y_i should be 1 when city i is selected, and 0 otherwise; x_{ij} should be 1 when city j is connected to selected city i, and 0 otherwise.

Lagrangian relaxation The method we are going to use to approximate the k-median problem is called Lagrangian relaxation. The idea is that we will modify the objective function, so that we pay a penalty if we select more than k cities. Let $\lambda \ge 0$. We now consider the following objective function: $\min \sum c_{ij} x_{ij} + \lambda(\sum y_i - k)$. Furthermore, we now remove the constraint that says we have to pick at most k cities. Clearly, any feasible solution to initial problem is also feasible here, for there are less constraints. Then, any feasible solution to the dual is still a lower bound to the original problem.

The dual to the Lagrangian relaxation is:

$$\max \sum \alpha_j - k \cdot \lambda$$

$$\text{subject to}$$

$$\sum_j \beta_{ij} \le \lambda, \text{ for all } i \in C$$

$$\alpha_j - \beta_{ij} \le c_{ij}, \text{ for all } i, j \in C$$

$$\beta \ge 0, \alpha \ge 0$$

Suppose that for some λ, the optimal solution of this problem has opened exactly k facilities. By looking at the term $\lambda \cdot (\sum y_i - k)$, we notice that it looks a lot like UFL with facility opening costs $f_i = \lambda$, minus the extra $k \cdot \lambda$. Then

$$\sum c_{ij} \overline{x}_{ij} \le 3 \left(\sum \alpha_j - \sum \lambda \overline{y}_i \right) = 3 \left(\sum \alpha - \lambda k \right) \le 3 \cdot OPT_{k-median}$$

the last inequality holding because $\sum \alpha_j - \lambda k$ is the cost of a dual feasible solution, which is a lower bound on $OPT_{k-median}$.

The discussion above assumes we know a λ such that the optimal solution opens k cities. How can we find such a λ? For $\lambda = 0$, we can open all cities at no cost. For λ sufficiently big, e.g. $n \cdot \max c_{ij}$, the optimum solution will open one facility only. The idea is that we can use *binary search* on integral values to try to find a good λ. If we manage to find a good λ, we are done. Otherwise, we will find λ_1 where the optimal solution opens k_1 cities, and λ_2 where the optimal solution opens k_2 cities, such that $k_1 < k < k_2$ and $(\lambda_1 - \lambda_2)$ is small (if we look only at integral λ's, the difference will be 1). Let (x^1, y^1) be a solution for λ_1 and (x^2, y^2) be a solution for λ_2. There exist a, b such that the fractional solution $a(x^1, y^1) + b(x^2, y^2)$ opens exactly k cities. We can do randomized rounding to take this fractional solution to an integral solution, and doing so we get a 6-approximation. This can be derandomized. An additional technique, called scaling, improves the analysis to obtain a 4-approximation.

All these results have been superceeded, in terms of approximation ratios and time bounds, but their development involved interesting ideas.

3 Dual Fitting Approach to UFL

To present the newer Greedy algorithms for UFL, we first examine the connection to Set Cover. In SC, we are given a family of subsets $\mathcal{S} = \{S_1, \ldots, S_m\}$ of the universe $U = \{u_1, \ldots, u_n\}$. There is a cost function c mapping $S \in \mathcal{S}$ to a non-negative real number c_S. The IP/LP formulation is

$$\min \sum_S c_S x_S$$

$$\text{subject to}$$

$$\sum_{S: u \in S} x_S \geq 1, \text{ for all } u \in U$$

Its dual is

$$\max \sum_{i \in U} y_i$$

$$\text{subject to}$$

$$\sum_{i \in S} y_i \leq c_S, \text{ for all } S \in \mathcal{S}$$

Below is the Greedy algorithm for SC.

while there exists $u \in U$ not covered
    find most cost effective not already selected set S % in terms of uncovered elements
    add this set S to the cover
    mark newly covered elements
end while

The price we pay for covering a certain element $e \in U$, *price*(e), is the average price per new element we pay the first time we cover e, i.e. the ratio $c_S / |U' \cap S|$ where S is the first set added to the cover that contains e, and U' is the set of uncovered elements just before S is added. It's easy to see that $\sum_{u \in U} price(u)$ is the cost of the cover. We have seen elsewhere that this cost is at most $H_d \cdot OPT$.

Before, the primal-dual approach could have been seen as a priority algorithm, in which whenever an equality becomes tight, we take that set. Now, we would like prices found by our Greedy algorithm to be a dual solution $y_i = price(i)$, in order to get a lower bound on the optimum. In fact, we claim that $y_i = price(i)/H_d$ is a dual feasible solution. This is not like the primal-dual algorithm seen before. Instead, we build a solution with a Greedy algorithm, which can be modified into a dual feasible solution, ultimately showing the lower bound on the optimum. The approach is called dual fitting.

UFL vs SC Let's return to the UFL problem. We know SC is a special case of UFL, where facilities are sets, cities are elements, and the cost function c_{ij} is ∞ when a element j is not in set i, and 0 otherwise. This cost function is not a metric except for some trivial cases.

Conversely, UFL can be seen as a SC problem. The sets (called stars) will be $\mathcal{S} = \{S_1, \ldots, S_m\}$, where $S_i = \langle i, C' \rangle$. For every subset of cities, $C' \subseteq C$, and every facility $i \in F$, we add a set of the form above. The elements of the universe in SC are the cities, and the set $\langle i, C' \rangle$ covers C' (i is not part of the universe, instead it's a notational convenience). The set $S_i = \langle i, C' \rangle$ has cost $c_{S_i} = cost(S_i) = \left(f_i + \sum_{j \in C'} c_{ij} \right) / |C'|$. Intuitively, picking the set $S_i = \langle i, C' \rangle$ in SC is equivalent in UFL to opening facility i and connecting cities in C' to i. Obviously, this is an exponential size system.

A Greedy algorithm is the most natural approach when we see UFL as SC. The Greedy algorithm we are going to use to solve this special instance of SC is based on the *most effective star*: at any point in time, we have selected some sets, which is equivalent to having opened certain facilities, and having connected certain cities to those facilities. At the next step, we perform the following:

for every facility i, opened or not
    sort the unconnected cities in non-decreasing connection cost to i
    suppose this list is $j_1, j_2, \ldots$
    look at the cost of stars of the form $\langle i, C' \rangle$ for $C' = \{j_1\}, \{j_1, j_2\}, \{j_1, j_2, j_3\}, \ldots$
    it can be shown that the cost of these stars initially decreases and then increases
    remember the most effective (least cost) star of this form for facility i
end for
after trying all facilities, return $\langle i', C' \rangle$, the most cost effective star

The following Greedy algorithm gives a constant factor approximation in the metric case, and a logarithmic factor in general.

initially, all cities are unconnected and all facilities are closed
let D be the set of unconnected cities
while D is not empty
    find most cost effective star $\langle i, C' \rangle$
    if i not already opened then
        open it
        set f_i to 0
    connect all $j \in C'$ to i
    $D \leftarrow D \setminus C'$

This is a priority algorithm, and we claim that if distances in the underlying UFL problem are metric, it gives a 1.86-approximation. It is similar to the Jain and Vazirany primal-dual algorithm. We have to make sure a city contributes to at most one facility. The analysis of this algorithm is by primal-dual fitting. We do not directly create a dual solution, but we can fit it to get a lower bound on the optimum.

Critical remark Consider applying the primal-dual strategy. If we start raising dual variables of all unconnected cities simultaneously, then the most cost-effective star $\langle i, C' \rangle$ will be the first pair such that $\sum_{j \in C'} \max(0, \alpha_j - c_{ij}) = f_i$. We can inductively correlate this approach with the Greedy algorithm.

Here, we do not allow more than one connection per city. At the end of the process, we do not obtain a dual solution. However, if we scale the α_j's down, then we get a dual feasible solution.

The equivalent Greedy algorithm in primal-dual terms is

$\alpha_j \leftarrow 0$ for all cities j
all facilities are closed
while C is not empty
    increase simultaneously α_j
    until one of the following occurs

(a) for some (or many) open i, $\alpha_j = c_{ij}$
(b) $\sum_{j \in C} \max(0, \alpha_j - c_{ij}) = f_i$

open i
connect every j with $\alpha_j \geq c_{ij}$ to i
remove all such j from C

As mentioned before, the α_j's obtained this way are not a dual feasible solution.

How do we analyze the approximation factor in this case? The H_d factor for Set Cover was relatively obvious, but this is much more involved. We need to set up another LP just to get the factor. This is called dual fitting with factor-revealing LP(IP).

References

Shmoys, Tardos and Aardal. *Approximation algorithms for facility location problems.* STOC 1997.

Jain and Vazirani. *Primal-dual approximation algorithms for metric facility location and k-median problems.* FOCS 1999.

Williamson. *The primal-dual method for approximation algorithms.* A survey.