An improved approximation algorithm for the facility location problem

Mohammad Mahdian

Based on joint works with Y. Ye and J. Zhang.

The Facility Location Problem

Uncapacitated Facility Location Problem (UFLP):

Given:

- set $\mathcal{F}$ of *facilities*,
- set $\mathcal{C}$ of *cities*,
- *opening cost* f_i for $i \in \mathcal{F}$, and
- *connection cost* c_{ij} for $i \in \mathcal{F}$ and $j \in \mathcal{C}$,

find:

- set $S \subseteq \mathcal{F}$ of facilities to *open*, and
- an assignment $\psi : \mathcal{C} \mapsto S$ of cities to *open* facilities

to minimize the total cost $\sum_{i \in S} f_i + \sum_{j \in \mathcal{C}} c_{\psi(j),j}$.

Assume the connection costs are metric.

JMS Algorithm

– Start at *time* $t = 0$. Set $\alpha_j = 0$ for every j.

– At any time, the amount that *unconnected* city j offers to contribute to facility i is $\max(\alpha_j - c_{ij}, 0)$. If j is connected to i', the amount of its offer to facility i is $\max(c_{i'j} - c_{ij}, 0)$.

– Increase α_j for all *unconnected* cities j at the same rate, until

- total amount offered to an *unopened* facility i equals f_i
  Open i and connect it to every city with a nonzero offer.
- for a city j and a facility i that is already open, $\alpha_j = c_{ij}$
  Connect j to i.

JMS Algorithm – Factor-Revealing LP

If γ is an upper bound on the solution of the following program

$$
\begin{aligned}
\text{maximize} \quad & \frac{\sum_{i=1}^{k} \alpha_i}{f + \sum_{i=1}^{k} d_i} \\
\text{subject to} \quad & \forall\, 1 \le i < k: \;\; \alpha_i \le \alpha_{i+1} \\
& \forall\, 1 \le j < i < k: \;\; r_{j,i} \ge r_{j,i+1} \\
& \forall\, 1 \le j < i \le k: \;\; \alpha_i \le r_{j,i} + d_i + d_j \\
& \forall\, 1 \le i \le k: \;\; \sum_{j=1}^{i-1} \max(r_{j,i} - d_j, 0) \\
& \qquad\qquad + \sum_{j=i}^{k} \max(\alpha_i - d_j, 0) \le f \\
& \forall\, 1 \le j \le i \le k: \;\; \alpha_j, d_j, f, r_{j,i} \ge 0
\end{aligned}
$$

then the JMS Algorithm is a γ-approximation!

More generally,

Definition. An algorithm is a (γ_f, γ_c)-approximation algorithm for FLP, if for every instance $\mathcal{I}$, and every solution SOL for $\mathcal{I}$ with facility cost F_{SOL} and connection cost C_{SOL}, the cost of the solution found by the algorithm is at most $\gamma_f F_{SOL} + \gamma_c C_{SOL}$.

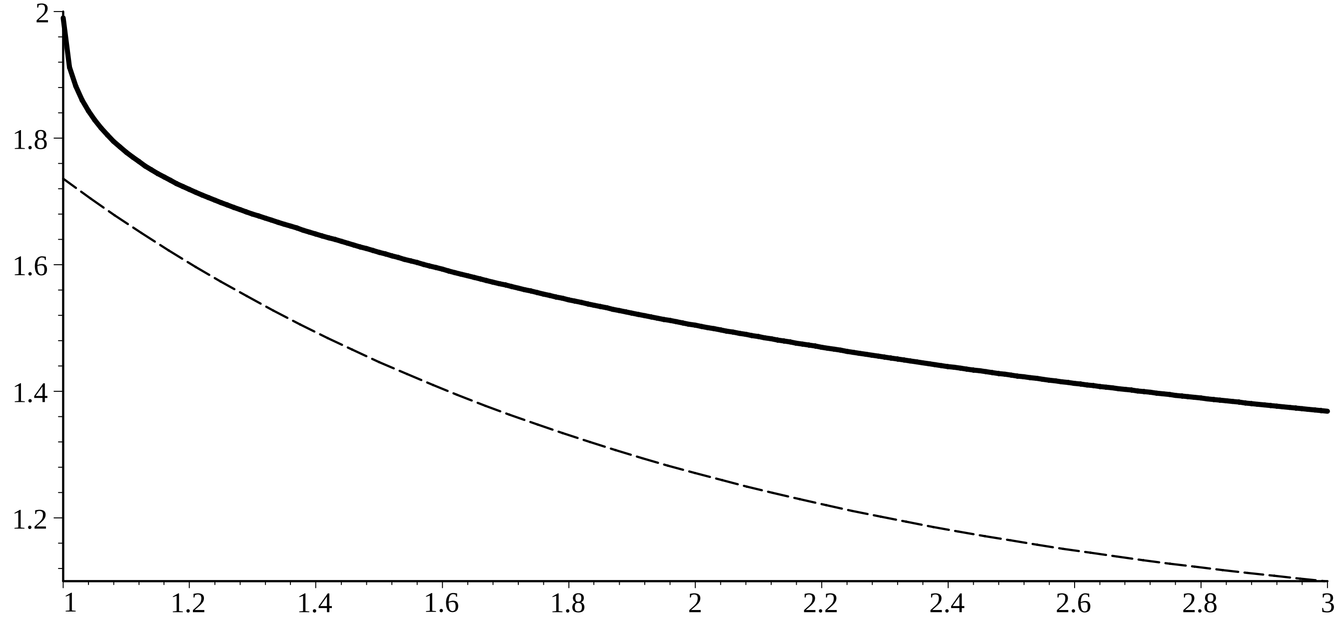
Theorem. The JMS algorithm is a (γ_f, γ_c)-approximation algorithm for UFLP, where for every γ_f, the value of γ_c can be computed using a similar factor-revealing LP.

A Tradeoff

For example,

The JMS Algorithm is a $(1.61, 1.61)$ -approximation algorithm.
$(1, 2)$
$(1.11, 1.78)$

In general, we can compute γ_c empirically:



Digression: Soft-capacitated FLP

It is usually useful to have separate approximation ratios for facility and connection costs.

Example:

Linear-cost FLP is similar to UFLP, except the cost of each open facility is a linear function of the number of cities it serves.

Theorem. There is a $(1, 2)$-approximation algorithm for LFLP.

Soft-capacitated FLP is similar to UFLP, except that each facility has a capacity u_i, and we are allowed to open multiple copies of each facility.

Theorem. There is a $(2, 1)$-*approximate reduction* from SCFLP to LFLP.

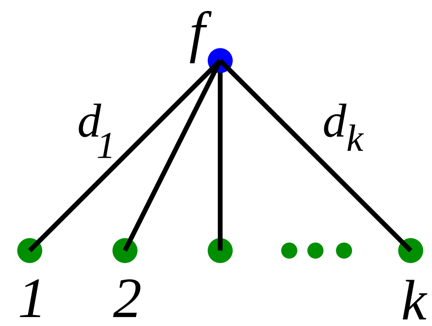
Corollary. There is a 2-approximation algorithm for SCFLP.

An improved algorithm for UFLP

1. Scale up the facility costs by a factor of δ.
2. Run the JMS Algorithm to solve the scaled-up instance.
3. Scale down the facility costs back to their original value, at the same rate. If at any point, a facility could be opened without increasing the total cost, open the facility and connect each city to the closest open facility.

   This step is equivalent to a greedy augmentation procedure of Charikar and Guha.

Consider a star (as in the analysis of the JMS Algorithm):



Variables $\alpha_j, d_j, f, r_{j,i}$ are defined as before (based on the first phase of the algorithm).

Analysis, cont'd

From the analysis of the JMS algorithm, we have the following inequalities:

$$\forall\, 1 \le i < k: \;\; \alpha_i \le \alpha_{i+1}$$

$$\forall\, 1 \le j < i < k: \;\; r_{j,i} \ge r_{j,i+1}$$

$$\forall\, 1 \le j < i \le k: \;\; \alpha_i \le r_{j,i} + d_i + d_j$$

$$\forall\, 1 \le i \le k: \;\; \sum_{j=1}^{i-1} \max(r_{j,i} - d_j, 0) + \sum_{j=i}^{k} \max(\alpha_i - d_j, 0) \le \delta f$$

$$\forall\, 1 \le j \le i \le k: \;\; \alpha_j, d_j, f, r_{j,i} \ge 0$$

Analysis, cont'd

In order to analyze the second phase, we slightly change it: Assume the scaling factor is changed discretely in L steps to

$$\delta = \delta_1 > \delta_2 > \cdots > \delta_L = 1.$$

After decreasing the scaling factor from δ_{i-1} to δ_i, we consider facilities in an arbitrary order, and open those that can be opened without increasing the total cost.

Let $r_{j,k+i}$ denote the connection cost that city j pays after we change the scaling factor to δ_i and process all facilities.

Therefore,

$$\sum_{j=1}^{k} \max(r_{j,k+i} - d_j, 0) \leq \delta_i f$$

Analysis, cont'd

How much does city j pay for facility costs in this algorithm?

In the first phase: α_j in total, $\alpha_j - r_{j,k+1}$ for facility cost.
After reducing the scaling factor to δ_{i+1}: $r_{j,k+i} - r_{j,k+i+1}$.

Therefore, total share of city j of facility costs is

$$\frac{\alpha_j - r_{j,k+1}}{\delta} + \sum_{i=1}^{L-1} \frac{r_{j,k+i} - r_{j,k+i+1}}{\delta_{i+1}}.$$

Thus, the total cost is:

$$\sum_{j=1}^{k} \left(\frac{\alpha_j - r_{j,k+1}}{\delta} + \sum_{i=1}^{L-1} \frac{r_{j,k+i} - r_{j,k+i+1}}{\delta_{i+1}} + r_{j,k+L+1} \right).$$

Analysis, cont'd

Therefore, the solution of the following factor-revealing LP gives an upper bound on the approximation ratio of the algorithm.

$$\text{maximize} \quad \frac{\sum_{j=1}^{k}\left(\frac{\alpha_j}{\delta}+\sum_{i=1}^{L-1}\left(\frac{1}{\delta_{i+1}}-\frac{1}{\delta_i}\right)r_{j,k+i}\right)}{f+\sum_{i=1}^{k} d_i}$$

$$\begin{aligned}
\text{subject to} \quad & \forall\, 1 \le i < k: \ \ \alpha_i \le \alpha_{i+1} \\
& \forall\, 1 \le j < i < k: \ \ r_{j,i} \ge r_{j,i+1} \\
& \forall\, 1 \le j < i \le k: \ \ \alpha_i \le r_{j,i} + d_i + d_j \\
& \forall\, 1 \le i \le k: \ \ \sum_{j=1}^{i-1} \max(r_{j,i} - d_j, 0) \\
& \qquad\qquad + \sum_{j=i}^{k} \max(\alpha_i - d_j, 0) \le \delta f \\
& \forall\, 1 \le i \le L: \ \ \sum_{j=1}^{k} \max(r_{j,k+i} - d_j, 0) \le \delta_i f \\
& \forall\, 1 \le j \le i \le k: \ \ \alpha_j, d_j, f, r_{j,i} \ge 0
\end{aligned}$$

Approximation ratio of the algorithm

Using the factor-revealing LP, it is easy to show that:

Theorem. If the JMS algorithm is a (γ_f, γ_c)-approximation algorithm for UFLP, then our algorithm is a $(\gamma_f + \ln(\delta), 1 + \frac{\gamma_c - 1}{\delta})$-approximation algorithm for UFLP.

Using $(\gamma_f, \gamma_c) = (1.11, 1.78)$ and $\delta = 1.504$, by the above theorem our algorithm is a 1.52-approximation algorithm for UFLP.

Open questions

- For every $\gamma_f \geq 1$, find a $(\gamma_f, 1 + 2e^{-\gamma_f})$-approx alg for UFLP.

  This will close the gap with the lower bound.

- Find a constant-factor approx alg for the *generalized FLP*.

The generalized FLP is similar to UFLP, except the cost of each facility is an arbitrary given function of the number of cities it serves.

Examples of special cases: UFLP, Soft-capacitated FLP, Hard-capacitated FLP, Load-balanced FLP