Chapter 4

The Uncapacitated Facility Location Problem

This chapter is based on two sources, namely, the survey by Cornuejols, Nemhauser and Wolsey [5, Chapter 3], and a recent paper by Aardal, Shmoys and Tardos [9]. The book by Nemhauser and Wolsey [6] has a detailed discussion on the topic of this chapter.

The results of Aardal et al are based on earlier results of Lin & Vitter [3, 4], and Raghavan & Thompson [8, 7].

4.1 The problem

Given a set of potential sites, a set of clients, and relevant profit and cost data, the goal is to find a maximum-profit plan giving the number of facilities to open, their locations and an allocation of each client to an open facility.

In detail, the input to an *uncapacitated facility location* (UFL) problem consists of

- a set $J = \{1, \ldots, n\}$ of potential sites for locating facilities,
- a set $I = \{1, \ldots, m\}$ of clients whose demands need to be served by the facilities,
- a profit c_{ij} for each $i \in I$ and $j \in J$; this is the profit made by satisfying the demand of client i from a facility located at site j, and
- a fixed nonnegative cost f_j for each $j \in J$; this is the (one time) cost of opening a facility at site j.

The problem is to select a subset Q ($Q \subseteq J$) of sites, to open facilities at these sites, and to assign each client to exactly one facility such that the difference of the variable profits and the fixed costs is maximized. The number of facilities to be opened, $|Q|$, is not prespecified, rather it is determined by an optimal solution. The profits c_{ij} usually depend on several factors such as the per unit production cost of a facility at site j, the per unit transportation cost from j to i, and the selling price to client i. An example problem and its optimal solution are shown in Figure 4.1.

$$\begin{aligned} m &= |I| = 3, \qquad n = |J| = 3 \\ f_j &= 1 \quad \text{for } (j = 1, \ldots, n) \\ C &= \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \end{aligned}$$

Optimal solution: $S = \{1, 2\}$, i.e., open facilities at sites 1 and 2. Assign client 1 to site 2, client 2 to site 1, and client 3 to site 1 to obtain an optimal value of 1.

Figure 4.1: An example of the uncapacitated facility location problem together with an optimal solution.

Proposition 4.1 *The UFL problem is NP-hard.*

Proof: Given an instance of the set covering problem, we can construct an instance of the UFL problem such that an optimal solution to the UFL problem gives an optimal solution to the set covering problem. First, we construct a bipartite graph based on the set covering instance: for each point x_i there is a node x_i in the "left side" of the bipartite graph, and for each set S_j, $S_j \subseteq \{x_1, x_2, \ldots, x_m\}$, there is a node S_j in the "right side" of the bipartite graph. There is an edge $x_i S_j$ if and only if point x_i belongs to set S_j. The instance of the UFL problem is as follows: the set of clients is $\{x_1, \ \ldots, \ x_m\}$, the set of potential sites is $\{S_1, S_2, \ldots, S_n\}$, the profit c_{ij} is taken to be either zero if edge $x_i S_j$ is present or $-\infty$ otherwise, and the fixed cost f_j (for each $j = 1, \ldots, n$) is taken to be 1.

Thus, the problem is to open the minimum number of facilities such that each client (point) x_i can be assigned to a facility (set) S_j adjacent to it (that contains x_i). It is easily seen that a solution of the UFL instance (set of sites to be opened) is optimal if and only if the corresponding solution of the set covering instance is optimal. ∎

4.2 Applications

The UFL problem is used to model many applications. Some of these applications are: bank account allocation, clustering analysis, lock-box location, location of offshore drilling platforms, economic lot sizing, machine scheduling and inventory management, portfolio management and the design of communication networks. We describe the first two applications. The *bank account location problem* arises from the fact that the clearing time for a check depends on the city i where it is cashed, and the city j where the paying bank is located. A company that pays bills by cheque to clients in several locations finds it useful to open accounts in several strategically located banks. It pays the bill to the client in city i from a bank in city j that maximizes the clearing time. Here I is the set of cities where clients are located, J is the set of potential bank locations, f_j is the cost of maintaining an account in city j, and c_{ij} is the monetary value of the clearing time between cities i and j. In *clustering analysis*, we are given a set I of objects, and the problem is to partition them into clusters such that objects in the same cluster are similar. Here, J is a subset of I, and consists of potential cluster representatives; the c_{ij}'s give the similarity between objects i and j; the f_j's may be either zero, in which case the clustering is entirely based on the similarity between objects, or the f_j's may be large, in which case the number of clusters tends to be minimized.

4.3 Linear programming formulations of the UFL problem

We start with the an integer linear programming formulation of the problem. For each potential site $j \in J$ we have a zero-one variable x_j. The intention is that a facility is opened at site j iff $x_j = 1$. For each client $i \in I$ and site $j \in J$, we have a zero-one variable y_{ij}. The intention is that the demand of client i is served by the facility at site j iff $y_{ij} = 1$.

$$
\begin{array}{llrcllr}
\text{(IP)} & \text{maximize} & z_{IP} & = & \sum_{i\in I}\sum_{j\in J} c_{ij}\, y_{ij} - \sum_{j\in J} f_j\, x_j & & \\
 & \text{subject to} & \sum_{j\in J} y_{ij} & = & 1 & \forall i \in I & \\
 & & y_{ij} & \leq & x_j & \forall i \in I, \forall j \in J & (I1) \\
 & & x_j & \in & \{0,1\} & \forall j \in J & \\
 & & y_{ij} & \in & \{0,1\} & \forall i \in I, \forall j \in J & (I2)
\end{array}
$$

Even if the integrality constraints on the y_{ij} are relaxed, i.e., even if the constraints (I2) above are replaced by

$$y_{ij} \geq 0 \qquad \forall i \in I,\ \forall j \in J,$$

the resulting mixed integer program is equivalent to (IP) above.
EXERCISE: Prove the above claim.

Another integer programming formulation, (WIP), is obtained from (IP) by replacing the mn constraints $(I1)$ by the n constraints

$$\sum_{i\in I} y_{ij} \leq m\, x_j \qquad \forall j \in J. \qquad (I1')$$

Each of the constraints $(I1')$ is obtained by summing the m constraints $y_{ij} \leq x_j$ in $(I1)$ for a fixed $j \in J$.
Fact: A set of numbers x_j $(j \in J)$, $y_{ij}(i \in I,\ j \in J)$ is a feasible solution of (WIP) if and only if it is a feasible solution of (IP).

Proof: Clearly, every feasible solution of (IP) is a feasible solution of (WIP). Consider a solution x_j, y_{ij} of (WIP). Suppose that for some $k \in J$, $x_k = 1$. Then the constraints $\sum_{j\in J} y_{ij} = 1$ (for each $i \in I$) ensure that each y_{ik} is at most 1. Otherwise, if $x_k = 0$, then $y_{ik} = 0$, $\forall i \in I$. Since all the constraints of (IP) are satisfied, the result follows. ∎

The linear programming relaxation of (IP) (respectively, (WIP)), called the strong (respectively, weak) LP relaxation and abbreviated (SLPR) (respectively, (WLPR)), is obtained by replacing the integrality restrictions on the x_j's by linear constraints, i.e., the constraints

$$x_j \in \{0,1\}, \qquad \forall j \in J$$

are replaced by the constraints

$$0 \leq x_j \leq 1, \qquad \forall j \in J.$$

$$\boxed{\begin{array}{llrcll} \text{(SLPR)} & \text{maximize} & z & = & \displaystyle\sum_{i\in I}\sum_{j\in J} c_{ij}\, y_{ij} \;-\; \sum_{j\in J} f_j\, x_j & \\ & \text{subject to} & \displaystyle\sum_{j\in J} y_{ij} & = & 1 & \forall i \in I \\ & & y_{ij} & \leq & x_j & \forall i \in I, \forall j \in J \\ & & x_j & \leq & 1 & \forall j \in J \\ & & x_j & \geq & 0 & \forall j \in J \\ & & y_{ij} & \geq & 0 & \forall i \in I, \forall j \in J \end{array}}$$

Fact: Every feasible solution of (SLPR) is a feasible solution of (WLPR), but there may be feasible solutions of (WLPR) that are not feasible solutions of (SLPR).

In practice, the strong LP relaxation performs remarkably well, giving integer optimal solutions on many instances of the UFL problem. The performance of (WLPR) is far poorer. Unfortunately, the computational cost of solving (SLPR) for large instances is high, since the number of constraints is $n + m + nm$. Specialized methods for solving (SLPR) are being developed by researchers.

4.4 Duality

To write down the dual of the strong LP relaxation, we introduce dual variables u_i, $i \in I$, w_{ij}, $i \in I$, $j \in J$, and t_j, $j \in J$, corresponding to the (SLPR) constraints $\sum_{j\in J} y_{ij} = 1$ $(\forall i \in I)$, $y_{ij} - x_j \leq 0$ $(\forall i \in I,\ \forall j \in J)$, and $x_j \leq 1$ $(\forall_j \in J)$, respectively. The dual LP is as follows:

$$\boxed{\begin{array}{llrcll} \text{(Dual SLPR)} & \text{minimize} & w & = & \displaystyle\sum_{i\in I} u_i + \sum_{j\in J} t_j & \\ & \text{subject to} & t_j - \displaystyle\sum_{i\in I} w_{ij} & \geq & -f_j & \forall j \in J \\ & & u_i + w_{ij} & \geq & c_{ij} & \forall i \in I,\ \forall j \in J \\ & & u_i & & \text{free} & \forall i \in I \\ & & w_{ij} & \geq & 0 & \forall i \in I,\ \forall j \in J \\ & & t_j & \geq & 0 & \forall j \in J \end{array}}$$

It is possible to write two condensed forms for (Dual SLPR), though the resulting duals are *not* linear programs.

4.4.1 First condensed dual

Suppose that all the variables u_i in (Dual SLPR) have fixed values. Then, to minimize the objective function, we must assign each w_{ij} the minimum value such that the constraints $u_i + w_{ij} \geq$

$c_{ij}\quad (\forall i \in I, \forall j \in J)$ and $w_{ij} \geq 0\quad (\forall i \in I, \forall j \in J)$ are satisfied. This gives

$$w_{ij} = (c_{ij} - u_i)^+, \qquad \forall i \in I, \forall j \in J,$$

where for an expression α, $(\alpha)^+$ means $\max(\alpha, 0)$. Now, consider the variables $t_j\quad (\forall j \in J)$. To minimize the objective function, we must assign each t_j the minimum value such that the constraints $t_j - \sum_{i\in I} w_{ij} \geq -f_j\quad (\forall j \in J)$ and $t_j \geq 0\quad (\forall j \in J)$ are satisfied. So, let

$$t_j = \left(\sum_{i\in I} w_{ij} - f_j\right)^+, \qquad \forall j \in J.$$

Substituting the formula for w_{ij} above we get

$$t_j = \left(\sum_{i\in I} (c_{ij} - u_i)^+ - f_j\right)^+, \qquad \forall j \in J.$$

This gives the first condensed dual:

$$\boxed{(\mathrm{CD1}) \qquad w = \underset{u_1,\ldots,u_m}{\text{minimize}} \left\{\textstyle\sum_{i\in I} u_i + \sum_{j\in J} \left(\sum_{i\in I}(c_{ij} - u_i)^+ - f_j\right)^+\right\}.}$$

4.4.2 Second condensed dual

In a feasible solution of the dual LP (Dual SLPR), suppose that there is a $k \in J$ such that t_k is positive, that is, $\left(\sum_{i\in I}(c_{ik} - u_i)^+ - f_k\right)^+$ is positive. Then there exists a $u_\ell (\ell \in I)$, such that $c_{\ell k} - u_\ell > 0$. If we increase u_ℓ by an amount $\epsilon(\epsilon > 0)$, then t_k will decrease by ϵ, therefore the objective value stays the same; also, all dual constraints will continue to hold. It follows that there always exists an optimal solution to the dual LP with

$$t_j \leq 0 \qquad \forall j \in J.$$

To see this, use the above procedure repeatedly until each t_j is at most zero. We also add the constraints $u_i \leq \max_{j\in J}(c_{ij})\quad (\forall i \in I)$; clearly, all optimal solutions of the dual LP satisfy these constraints. This gives the second condensed dual:

$$\boxed{\begin{array}{llcll}
(\mathrm{CD2}) & \text{minimize} & w & = & \sum_{i\in I} u_i & \\
 & \text{subject to} & \sum_{i\in I}(c_{ij} - u_i)^+ - f_j & \leq & 0 & \forall j \in J \\
 & & u_i & \leq & \max_{j\in J}(c_{ij}) & \forall i \in I.
\end{array}}$$

4.5 Heuristics for solving the UFL problem

This section develops two well known heuristics for solving the UFL problem, namely, the greedy heuristic and the dual descent procedure. These heuristics simultaneously find a candidate solution $x_j \in \{0,1\}, y_{ij} \in \{0,1\} \quad (\forall j \in J,\ \forall i \in I)$ to the UFL problem instance as well as a feasible solution for the dual of the strong LP relaxation. By linear programming weak duality, the objective value of every feasible solution of (DUAL SLPR) gives an upper bound on the optimal value of (SLPR), and hence it gives an upper bound on the optimal value of the UFL instance. So, not only do we obtain a candidate solution to the UFL instance, but also an indication of how "far" this solution is from being optimal.

4.5.1 The greedy heuristic

We start with an empty set S of open facilities, and at each step we add to S a site $j \in J \backslash S$ that yields the maximum improvement in the objective value. For a set S, $S \subseteq J$, of open facilities, the objective value is given by

$$z(S) = \sum_{i \in I} \max_{j \in S}\{c_{ij}\} - \sum_{j \in S} f_j.$$

For a site $j \in J \backslash S$, let $p_j(S) = z(S \cup \{j\}) - z(S)$ denote the change in the objective value when a new facility is opened at j. For the currently open set of facilities S and for each client $i \in I$, define $u_i(S)$ to be $\max_{j \in S}\{c_{ij}\}$; define $u_i(\emptyset)$ to be 0. That is, $u_i(S)$ is the maximum profit obtained from serving client i using only the facilities in S. Then

$$z(S) = \sum_{i \in I} u_i(S) - \sum_{j \in S} f_j$$

and further

$$p_j(S) = z(S \cup \{j\}) - z(S) = \sum_{i \in I} (c_{ij} - u_i(S))^+ - f_j.$$

In each iteration of the greedy heuristic, we compute $p_j(S)$ for each $j \in J \backslash S$. If either $J \backslash S$ is empty or $p_j(S) \leq 0$ for each $j \in J \backslash S$, then we terminate the heuristic. Otherwise, we add to S a $j \in J \backslash S$ whose incremental value $p_j(S)$ is maximum. See Figure 4.2 for an example.

Consider the first condensed dual of (SLPR), (CD1), and for each $i \in I$ let the ith dual variable u_i be assigned the value $u_i(S)$ defined above. Then the dual objective value corresponding to S is

$$\begin{aligned} w(u(S)) &= \sum_{i \in I} u_i(S) + \sum_{j \in J} \left(\sum_{i \in I} (c_{ij} - u_i(S))^+ - f_j \right)^+ \\ &= \sum_{i \in I} u_i(S) + \sum_{j \in S} (p_j(S))^+ , \end{aligned}$$

because each $j \in S$ has $\sum_{i \in I} (c_{ij} - u_i(S))^+ - f_j = p_j(S)$.

For the final solution S^G found by the greedy heuristic, each $j \in J \backslash S^G$ has $p_j(S^G) \leq 0$, hence the dual objective value is

$$w(u(S^G)) = \sum_{i \in I} u_i(S^G).$$

Figure 4.2: Solving an example of the uncapacitated facility location problem by the greedy heuristic (from Chapter 3, "Discrete Location Theory", Ed., Mirchandani & Francis).

$$\begin{aligned} m &= 4, \quad n = 6 \\ f &= [\ 3 \ \ 2 \ \ 2 \ \ 2 \ \ 3 \ \ 3\ ] \\ C &= \begin{bmatrix} 6 & 6 & 8 & 6 & 0 & 6 \\ 6 & 8 & 6 & 0 & 6 & 6 \\ 5 & 0 & 3 & 6 & 3 & 0 \\ 2 & 3 & 0 & 2 & 4 & 4 \end{bmatrix}. \end{aligned}$$

Iteration 1:

$$S_0 = \emptyset, \quad Z(S_0) = 0, \quad u(S_0) = [0\ 0\ 0\ 0].$$

$p_1(S_0)$	$p_2(S_0)$	$p_3(S_0)$	$p_4(S_0)$	$p_5(S_0)$	$p_6(S_0)$
16	15	15	12	10	13

$$w(u(S_0)) = (0) + (81) = 81.$$
$$S_1 := \{1\}, \ \text{ since } p_1(S_0) > 0 \text{ is maximum.}$$

Iteration 2:

$$S_1 = \{1\}, \quad z(S_1) = 16, \quad u(S_1) = [6\ 6\ 5\ 2].$$

$p_2(S_1)$	$p_3(S_1)$	$p_4(S_1)$	$p_5(S_1)$	$p_6(S_1)$
$(2+1)-2=1$	$(2)-2=0$	$(1)-2=-1$	$(2)-3=-1$	$(2)-3=-1$

$$w(u(S_1)) = (19) + (1) = 20.$$
$$S_2 := \{1,2\}, \ \text{ since } p_2(S_1) > 0 \text{ is maximum.}$$

Iteration 3:

$$S_2 = \{1,2\}, \quad z(S_2) = 16 + 1 = 17, \quad u(S_2) = [6\ 8\ 5\ 3].$$

$p_3(S_2)$	$p_4(S_2)$	$p_5(S_2)$	$p_6(S_2)$
$(2)-2=0$	$(1)-2=-1$	$(1)-3=-2$	$(1)-3=-2$

$$w(u(S_2)) = (22) + (0) = 22.$$
$$\textbf{stop}, \text{ since } \ p_j(S_2) \leq 0, \quad \forall j \in J \backslash S_2.$$

The greedy solution is $S^G = \{1,2\}$, with objective value $Z^G = 17$.
The dual greedy value W^G is the best upper bound computed on the optimal value: $W^G = \min_{t=1,2,\ldots} w(u(S_t)) = w(u(S_1)) = 20$

The greedy heuristic actually computes the dual objective value $w(u(S))$ for the set S in each iteration, and takes the smallest of these values to be the upper bound W^G that is returned at the end, since this value is the least upper bound computed on the optimal value of (SLPR). See the example in Figure 4.2.

The proofs of the next two results may be found in Cornuejols, Fisher and Nemhauser (1977). Below, $e \approx 2.71828$ denotes the base of the natural logarithm.

Theorem 4.2 *For the UFL problem, the objective value Z^G of the candidate solution found by the greedy heuristic is at least*

$$\frac{e-1}{e}W^G + \frac{1}{e}\left(\sum_{i\in I}\min_{j\in J} c_{ij} - \sum_{j\in J} f_j\right).$$

Theorem 4.3 *For the k-median problem with $c_{ij} \geq 0$ for all i and j (here, $f_j = 0$ for all j), the objective value Z^G of the candidate solution found by the greedy heuristic is at least*

$$\frac{e-1}{e}W^G.$$

4.5.2 The dual descent procedure

The dual descent procedure for solving the UFL problem works well on most instances, but it may perform poorly on hard instances. This procedure is used by the program DUALOC (by Erlenkotter (1978)), which is one of the best programs available for solving UFL problems.

The procedure attempts to find a good solution $u_1, \ldots, u_m$ to the second condensed dual, (CD2), and then uses the complementary slackness conditions to find a candidate solution for the UFL problem instance. The dual variables u_i $(\forall i \in I)$ are initialized to $\max_{j\in J}\{c_{ij}\}$. This gives a feasible solution to the constraints

$$\sum_{i\in I}(c_{ij} - u_i)^+ - f_j \leq 0 \qquad \forall j \in J \qquad (*)$$

of (CD2). Then the procedure repeatedly steps through all the indices $i \in I$ in an arbitrary but fixed order and attempts to decrease u_i as follows: if u_i can be decreased to

$$\max_{j\in J}\ \{c_{ij} : c_{ij} < u_i\}$$

(i.e., the largest profit c_{ij} for client i that is strictly less than u_i) without violating the constraint $(*)$ above, then this is done, otherwise u_i is decreased to the smallest value such that the constraint $(*)$ continues to be satisfied. The procedure terminates when there is an iteration such that no u_i $(\forall i \in I)$ is decreased. The complementary slackness conditions for finding a candidate solution to the UFL instance are as follows: A pair of feasible solutions x_j, y_{ij} $(\forall i \in I,\ \forall j \in J)$ and u_i $(\forall i \in I)$ to (SLPR) and it dual, respectively, forms optimal solutions only if

$$x_j = 0 \qquad \text{or} \qquad t_j - \sum_{i\in I} w_{ij} = -f_j \qquad (\forall j \in J),$$

or, in term of (CD2), only if

$$x_j = 0 \qquad \text{or} \qquad \sum_{i \in I}(c_{ij} - u_i)^+ - f_j = 0 \qquad (\forall j \in J).$$

This gives us the following complementarity conditions:

$$x_j \left(\sum_{i \in I}(c_{ij} - u_i)^+ - f_j \right) = 0 \qquad \forall j \in J.$$

Given a feasible solution $(u_1, \ldots, u_m)$ of (CD2), if $\sum_{i \in I}(c_{ij} - u_i)^+ - f_j < 0$, then we fix x_j at zero. Let $J(u)$, $J(u) \subseteq J$, be the set of all sites j such that the constraint $(*)$ holds with equality, i.e.,

$$J(u) = \left\{ j \in J : \sum_{i \in I}(c_{ij} - u_i)^+ - f_j = 0 \right\}.$$

Then, we find a minimal subset $K(u)$ of $J(u)$ such that for all $i \in I$, $\max_{j \in K(u)} \{c_{ij}\} = \max_{j \in J(u)} c_{ij}$ (possibly, $K(u) = J(u)$). The set $K(u)$ is the set of sites where facilities are opened. The objective value of the UFL instance corresponding to $K(u)$ is given by

$$Z^{DD} = \sum_{i \in I} \max_{j \in K(u)} \{c_{ij}\} - \sum_{j \in K(u)} f_j.$$

An example is shown in Figure 4.3.

The dual descent procedure cannot find the optimal solution to every instance of the UFL problem, simply because the optimal value of the UFL instance may be less than the optimal value of the strong LP relaxation. Moreover, the procedure is *not* guaranteed to find an optimal solution to the condensed dual (CD2). The example in Figure 4.4 has an optimal dual solution $u^* = (1, 1, 1)$ with objective value $w^* = 3$, however, the solution found by the dual descent procedure is $u^{DD} = (0, 2, 2)$ with objective value $w^{DD} = 4$.

Proposition 4.4 *Let $K(u)$ be defined as above. For each client i, let k_i be the number of sites in $K(u)$ whose profit c_{ij} is greater than u_i, $k_i = |\{j \in K(u) : c_{ij} > u_i\}|$. If k_i is at most one for each $i \in I$, then $K(u)$ is an optimal set of open facilities for the UFL problem.*

Proof: Focus on a client i. If $k_i = 0$, then we have

$$\max_{j \in K(u)} \{c_{ij}\} = u_i = u_i + \sum_{j \in K(u)} (c_{ij} - u_i)^+,$$

and if $k_i = 1$, then we have

$$\max_{j \in K(u)} \{c_{ij}\} = u_i + \sum_{j \in K(u)} (c_{ij} - u_i)^+.$$

Figure 4.3: Solving an example of the uncapacitated facility location problem by the dual descent procedure (from Chapter 3, “Discrete Location Theory”, Ed., Mirchandani & Francis).

$$\begin{aligned} m &= 4, \quad n = 6 \\ f &= [\ 3 \ \ 2 \ \ 2 \ \ 2 \ \ 3 \ \ 3\ ] \\ C &= \begin{bmatrix} 6 & 6 & 8 & 6 & 0 & 6 \\ 6 & 8 & 6 & 0 & 6 & 6 \\ 5 & 0 & 3 & 6 & 3 & 0 \\ 2 & 3 & 0 & 2 & 4 & 4 \end{bmatrix}. \end{aligned}$$

Step no.	u_1	u_2	u_3	u_4	$\sum_{i\in I}(c_{ij}-u_i)^+ - f_j$ $j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$
0	8	8	6	4	-3	-2	-2	-2	-3	-3
1	6	8	6	4	-3	-2	0	-2	-3	-3
2	6	6	6	4	-3	0	0	-2	-3	-3
3	6	6	5	4	-3	0	0	-1	-3	-3
4	6	6	5	3	-3	0	0	-1	-2	-2
5	6	6	4	3	-2	0	0	0	-2	-2

In the last iteration, u_1 cannot decrease otherwise the constraint

$$\sum_{i\in I}(c_{ij}-u_i)^+ - f_j \leq 0 \qquad (*)$$

is violated for $j = 3$, and u_2 cannot decrease otherwise the constraint $(*)$ for $j = 2$ is violated; u_3 is decreased from 5 to 4 (not 3, since that violates the constraint $(*)$ for $j = 4$); u_4 cannot decrease otherwise the constraint $(*)$ for $j = 2$ is violated.

The dual objective value is $w^{DD}(u) = \sum_{i\in I} u_i = 19$, $J(u) = \{2, 3, 4\}$, and $K(u) = \{2, 3, 4\}$.

Opening facilities at sites 2, 3, and 4 gives an objective value of $Z^{DD} = (8+8+6+3)-(2+2+2) = 19$. This solution is optimal, since $Z^{DD} = w^{DD}(u)$. Optimality also follows from Proposition 4.4.

Figure 4.4: An example of the uncapacitated facility location problem on which the dual descent procedure fails to find an optimal solution for the dual (CD2) (from Chapter 3, "Discrete Location Theory", Ed., Mirchandani & Francis).

$$\begin{aligned} m &= 3, \quad n = 3 \\ f &= [\ 2 \ \ 2 \ \ 2\ ] \\ C &= \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix}. \end{aligned}$$

Step no.	u_1	u_2	u_3	$\sum_{i \in I} (c_{ij} - u_i)^+ - f_j$, $j = 1$	$j = 2$	$j = 3$
0	2	2	2	−2	−2	−2
1	0	2	2	−2	0	0

In the last iteration, u_1, u_2 and u_3 are prevented from decreasing further by the constraints $(*)$ for $j = 2$, 3 and 2, respectively.

$$w^{DD}(u) = \sum_{i \in I} u_i = 4 > w^* = 3.$$

Now, consider the objective value of the UFL instance corresponding to the given solution $K(u)$,

$$\begin{aligned} z(K(u)) &= \sum_{i \in I} \max_{j \in K(u)} \{c_{ij}\} - \sum_{j \in K(u)} f_j \\ &= \sum_{i \in I} \left(u_i + \sum_{j \in K(u)} (c_{ij} - u_i)^+ \right) - \sum_{j \in K(u)} f_j, \\ &= \sum_{i \in I} u_i + \sum_{j \in K(u)} \left(\sum_{i \in I} (c_{ij} - u_i)^+ - f_j \right) \\ &= \sum_{i \in I} u_i, \quad \text{(since complementarity ensures that for each } j \in K(u), \\ &\qquad\qquad \sum_{i \in I} (c_{ij} - u_i)^+ - f_j \quad \text{is at zero)} \\ &= w(u). \end{aligned}$$

Since the dual objective value equals $z(K(u))$, $K(u)$ must be an optimal solution. ∎

4.6 The filtering and rounding method for location problems

4.6.1 Introduction

This section describes the filtering and rounding method, a recently developed method for finding approximately optimal solutions to NP-hard location problems. The method starts from an optimal (or near-optimal) solution to an appropriately formulated LP (linear programming) relaxation, and obtains a near-optimal solution to the original location problem. There are two major steps. The filtering step uses the optimal LP solution to construct a "filtered problem" by fixing some variables to zero in the integer programming formulation. Thus the filtered problem is a restricted version of the original location problem. The critical point is this:

> *every integral solution* of the filtered problem is guaranteed to have its objective value "near" (i.e., within a $(1+\epsilon)$-factor of) the optimal value of the original LP.

The rounding step produces a feasible integral solution to the filtered problem, either by using a simple randomized heuristic for "rounding" a "fractional solution" (i.e., a solution to the LP relaxation) to the filtered problem, or by using a simple greedy heuristic (nonrandomized) for the same task.

In our application, the "minimize" UFL problem with symmetric costs that satisfy the triangle inequality, the method does give a feasible integral solution whose objective value is "near" the optimal value.

The filtering and rounding method originated from the work of Lin & Vitter (1992). They apply it to the k-median problem. The application to the "minimize" UFL problem with restricted costs is due to Aardal, Shmoys and Tardos (1997).

4.6.2 An integer programming formulation for "minimize" UFL problems and its LP relaxation

Let V be a set of nodes. Let $I \subseteq V$ be a set of *client nodes*, and let $J \subseteq V$ be a set of *site nodes*. For each pair of nodes $v, w \in V$, let c_{vw} be the *service cost* for v, w. For each site node $j \in J$, let f_j be the *fixed cost* for opening a facility at j. The "minimize" UFL problem is to open facilities at a subset of sites $J^* \subseteq J$ and to assign each client $i \in I$ to a facility $j \in J^*$ so as to minimize the sum of the "service costs", namely, $\sum_{i \in I} \{c_{ij} \mid j \in J^* \text{ is the unique facility assigned to serve client } i\}$, and the "fixed costs", namely, $\sum_{j \in J^*} f_j$.

Here is an integer linear programming formulation of the "minimize" UFL problem. Let (IP) denote this integer linear program, and let z_{IP} denote the optimal value of (IP). For each site $j \in J$ we have a zero-one variable y_j. The intention is that a facility is opened at site j iff $y_j = 1$. For each client $i \in I$ and site $j \in J$, we have a zero-one variable x_{ij}. The intention is that the demand of client i is served by the facility at site j iff $x_{ij} = 1$.

$$
\boxed{
\begin{array}{llrcll}
\text{(IP)} & z_{IP} & & = & \text{minimize } \sum_{i\in I}\sum_{j\in J} c_{ij}\, x_{ij} + \sum_{j\in J} f_j\, y_j & \\
& \text{subject to} & \sum_{j\in J} x_{ij} & = & 1 & \forall i \in I \\
& & x_{ij} & \le & y_j & \forall i \in I, \forall j \in J \\
& & y_j & \in & \{0,1\} & \forall j \in J \\
& & x_{ij} & \in & \{0,1\} & \forall i \in I, \forall j \in J
\end{array}
}
$$

The linear programming relaxation of (IP) is obtained by replacing the integrality restrictions on the y_j's by linear constraints, i.e., the constraints

$$y_j \in \{0,1\}, \qquad \forall j \in J$$

are replaced by the constraints

$$0 \le y_j \le 1, \qquad \forall j \in J.$$

Also, the integrality restrictions on the x_{ij}'s are replaced by nonnegativity constraints

$$0 \le x_{ij}, \qquad \forall i \in I, \forall j \in J.$$

Let (LP) denote this relaxation, and let z_{LP} denote the optimal value of (LP).

4.6.3 The Aardal-Shmoys-Tardos algorithm for "minimize" UFL problems

Recall that a function $c : V \times V \to \Re$ is said to satisfy the *triangle inequality* if the following holds

$$\text{for all triples of nodes } u, v, w, \quad c(v,w) \le c(v,u) + c(u,w).$$

Theorem 4.5 (Aardal, Shmoys & Tardos (1997)) *Suppose that the service costs c and the fixed costs f are nonnegative. If the costs matrix $[c_{vw}]$ $(\forall v, w \in V)$ is symmetric and satisfies the triangle inequality, then there is a 3-approximation algorithm for the "minimize" UFL problem. In fact, the approximation algorithm delivers a feasible integral solution $\widetilde{x}, \widetilde{y}$ to (IP) whose objective value is at most $3z_{LP}$.*

We will prove a weaker version of this theorem. The proof will describe an approximation algorithm, and we will prove that it achieves an approximation guarantee of 4. Let ϵ be a fixed parameter. (It will turn out that the best choice is $\epsilon = 1/3$.) Focus on the objective function of (IP) or (LP), and note that it consists of a "service cost" part $(\sum_{i\in I}\sum_{j\in J} c_{ij}x_{ij})$ and a "fixed cost" part $(\sum_{j\in J} f_j y_j)$. Let z^s_{LP} and z^f_{LP} denote the values of the "service cost" part and of the "fixed cost" part in an optimal solution to (LP).

Let $\widehat{x}, \widehat{y}$ be an optimal solution to (LP). For each client $i \in I$, define $\widehat{C}_i = \sum_{j\in J} c_{ij}\widehat{x}_{ij}$. Note that $z^s_{LP} = \sum_{i\in I}\widehat{C}_i$. As in the Lin & Vitter algorithm for the k-median problem, take $V_i = \{j \in J \mid c_{ij} \le (1+\epsilon)\widehat{C}_i\}$.

Recall the main lemma in Lin & Vitter's analysis of their algorithm.

Lemma 4.6 *For all $i \in I$,*

$$\sum_{j \in V_i} \widehat{y}_j > \frac{\epsilon}{1+\epsilon}.$$

We construct a "near optimal" feasible solution $\widetilde{x}, \widetilde{y}$ to the "filtered" (LP), by taking

$$\widetilde{y}_j = \min\{1,\ (\frac{1+\epsilon}{\epsilon})\widehat{y}_j\}, \quad \forall j \in J.$$

By the above lemma, for all $i \in I$, we have $\sum_{j \in V_i} \widetilde{y}_j \geq 1$.

We take $[\widetilde{x}_{ij}]$ $(\forall i \in I, \forall j \in J)$ to be optimal with respect to $\widetilde{y}_j$ $(\forall j \in J)$. (That is, we take $\widetilde{y}$ to be fixed, and use the greedy procedure to find $\widetilde{x}$ such that the service costs are minimized.)

The second part of the algorithm starts with the "fractional" solution $\widetilde{x}, \widetilde{y}$ and executes several iterations. In each iteration, at least two fractional variables $\widetilde{y}_j$ are assigned integer values. In more detail, one fractional variable $\widetilde{y}_j$ is assigned the value 1, and one or more fractional variables $\widetilde{y}_j$ are assigned the value 0. This immediately implies that several of the fractional variables $\widetilde{x}_{ij}$ are assigned integer values of 0 or 1. We use $\widetilde{x}, \widetilde{y}$ to denote the current solution, through all iterations. At the start of the second part, $\widetilde{x}, \widetilde{y}$ is obtained from the optimal LP solution as described above. At termination, $\widetilde{x}, \widetilde{y}$ will be a feasible integral solution to (IP), and its objective value will be proved to be $\leq 4z_{LP}$.

Let $\widetilde{J} = \{j \in J \mid 0 < \widetilde{y}_j < 1\}$ denote the set of sites whose variables have fractional values in the current solution $\widetilde{y}$. Let $\widetilde{I} = \{i \in I \mid \exists j \in J \text{ such that } 0 < \widetilde{x}_{ij} < 1\}$ denote the set of clients that are partially assigned to a "fractional site."

Fact 4.7 *For each client i in $\widetilde{I}$, there is no site $j \in J$ such that $\widetilde{x}_{ij} > 0$ and $\widetilde{y}_j = 1$, and moreover, $\{j \in J \mid \widetilde{x}_{ij} > 0\}$ must contain at least two "fractional sites" $j \in \widetilde{J}$.*

The goal in each iteration is to maintain the following:

Induction hypothesis:

(1) $\widetilde{x}, \widetilde{y}$ is a feasible solution to (LP).

(2) $\displaystyle\sum_{j \in J} f_j \widetilde{y}_j \quad \leq \quad (\frac{1+\epsilon}{\epsilon}) z_{LP}^f \quad = \quad (\frac{1+\epsilon}{\epsilon}) \sum_{j \in J} f_j \widehat{y}_j.$

(3) $\forall i \in \widetilde{I}\ :\ \widetilde{x}_{ij} > 0 \Longrightarrow c_{ij} \leq (1+\epsilon)\widehat{C}_i.$

(4) $\forall i \in I\ :\ \displaystyle\sum_{j \in J} c_{ij} \widetilde{x}_{ij} \leq 3(1+\epsilon)\widehat{C}_i.$

The induction basis is that parts (1)–(4) of the induction hypothesis hold at the start (of the second part of the algorithm). Part (1) follows from Lin & Vitter's main lemma; part (2) holds by our choice of $\widetilde{y}_j \leq (1+\epsilon)\widehat{y}_j/\epsilon$, $\forall j \in J$; parts (3) and (4) hold by our choice of V_i, $\forall i \in I$.

An iteration starts by choosing a client $h \in \widetilde{I}$ that minimizes $\widehat{C}_h$, i.e., $\widehat{C}_h \leq \widehat{C}_i, \forall i \in \widetilde{I}$.

Let $L = \{j \in J \mid \widetilde{x}_{hj} > 0\}$, and note that $L \subseteq \widetilde{J}$. That is, L consists of the sites assigned to serve the client h by the current solution $\widetilde{x}, \widetilde{y}$. By the fact above, every site in L is "fractional", and L must have ≥ 2 sites.

Let $H = \{i \in I \mid \exists j \in L \text{ such that } \widetilde{x}_{ij} > 0\}$, and note that $H \subseteq \widetilde{I}$. That is, H consists of those clients who are served by at least one "fractional site" in L according to the current solution $\widetilde{x}, \widetilde{y}$.

Choose $\ell \in L$ that minimizes f_ℓ, i.e., $f_\ell \leq f_j, \forall j \in L$. In words, ℓ is the "fractional site" serving our chosen client h that has the smallest fixed cost.

> **We now change the current solution $\widetilde{x}, \widetilde{y}$ to $\widetilde{x}', \widetilde{y}'$:**
> we take $\widetilde{y}'_\ell = 1$ and take $\widetilde{y}'_j = 0, \forall j \in L - \{\ell\}$;
> for all clients $i \in H$, we take $\widetilde{x}'_{i\ell} = 1$ and take $\widetilde{x}'_{ij} = 0, \forall j \in J - \{\ell\}$;
> for all the remaining variables, we take $\widetilde{y}'_j = \widetilde{y}_j$ and $\widetilde{x}'_{ij} = \widetilde{x}_{ij}$.

Finally, we must prove that the new solution $\widetilde{x}', \widetilde{y}'$ satisfies the induction hypothesis, assuming that the previous solution $\widetilde{x}, \widetilde{y}$ did so.

Part(1) obviously holds.

For part(2), note that $\sum_{j \in J} f_j \widetilde{y}_j = \sum_{j \in L} f_j \widetilde{y}_j + \sum_{j \in J-L} f_j \widetilde{y}_j$. Focus on the first term, $\sum_{j \in L} f_j \widetilde{y}_j$. We claim that it is at least $f_\ell = \sum_{j \in L} f_j \widetilde{y}'_j$. To see this, firstly note that $\sum_{j \in L} \widetilde{y}_j \geq 1$ (since $\widetilde{x}, \widetilde{y}$ is feasible for (LP), we have $\sum_{j \in J} \widetilde{x}_{hj} = \sum_{j \in L} \widetilde{x}_{hj} = 1$ and $\widetilde{x}_{hj} \leq \widetilde{y}_j, \forall j \in J$). Secondly, note that $f_\ell \leq f_j, \forall j \in L$, by our choice of ℓ.

Part(3) obviously holds because all clients i in the new $\widetilde{I}$ were in $\widetilde{I} - H$, and so have the same values before and after the iteration for their variables $\widetilde{x}_{ij} (\forall j \in J)$.

For part(4), consider any client $i \in H$. We claim that $c_{i\ell} \leq 3(1+\epsilon)\widehat{C}_i$. To see this, firstly note that by the definition of H, there must be a site $j \in L$ with $\widetilde{x}_{ij} > 0$, hence $c_{ij} \leq (1+\epsilon)\widehat{C}_i$. Secondly, note that by the definition of L, both $\widetilde{x}_{hj} > 0$ and $\widetilde{x}_{h\ell} > 0$, hence $c_{hj} \leq (1+\epsilon)\widehat{C}_h$ and $c_{h\ell} \leq (1+\epsilon)\widehat{C}_h$. Since $[c_{vw}]$ is symmetric and satisfies the triangle inequality, and moreover, we chose client h to minimize $\widehat{C}_h$ over all $i \in \widetilde{I}$, we have

$$c_{i\ell} \leq c_{ij} + c_{jh} + c_{h\ell} \leq c_{ij} + c_{hj} + c_{h\ell} \leq (1+\epsilon)(\widehat{C}_i + \widehat{C}_h + \widehat{C}_h) \leq (1+\epsilon)(3\widehat{C}_i).$$

To derive the approximation guarantee of 4, we take $\epsilon = 1/3$ and apply the induction hypothesis to the final solution $\widetilde{x}, \widetilde{y}$, which is guaranteed to be integral. Then we see that the final solution $\widetilde{x}, \widetilde{y}$ has service cost

$$\leq 3(1+\epsilon)\sum_{i \in I} \widehat{C}_i = 3(1+\epsilon) z^s_{LP} \leq 4 z^s_{LP},$$

(by part(3) of the induction hypothesis), and has fixed cost

$$\leq \left(\frac{1+\epsilon}{\epsilon}\right) z^f_{LP} \leq 4 z^f_{LP},$$

(by part(4) of the induction hypothesis).

4.7 Exercises

1. (From Chapter 3, "Discrete Location Theory", Ed., Mirchandani and Francis.) Consider the instance of the UFL problem defined by $m = 5,\ n = 8,\ f_j = 2$ for $j = 1, \ldots, 8$, and

$$C = \begin{bmatrix} 3 & 6 & 3 & 5 & 6 & 4 & 3 & 0 \\ 4 & 4 & 5 & 3 & 5 & 2 & 0 & 5 \\ 4 & 3 & 4 & 3 & 4 & 0 & 4 & 5 \\ 5 & 3 & 4 & 6 & 0 & 4 & 6 & 3 \\ 5 & 5 & 4 & 0 & 3 & 6 & 5 & 3 \end{bmatrix}.$$

   (a) Use the greedy heuristic to find a solution. Give the greedy value Z^G and the dual greedy value W^G.

   (b) Give the optimal value W' of the weak linear programming relaxation (use the CPLEX optimizer).

   (c) Apply the dual descent procedure, cycling through the indices i. Give the value $W^{DD}(u)$ found by this procedure. Is the set $K(u)$ an optimal set of open facilities for this problem instance?

2. Repeat parts (a)–(c) of the previous problem for another instance of the UFL problem defined by $m = 4,\ n = 5,\ f = [1\ 2\ 3\ 4\ 5]$, and

$$C = \begin{bmatrix} 9 & 7 & 6 & 7 & 8 \\ 7 & 9 & 8 & 5 & 6 \\ 8 & 7 & 10 & 9 & 8 \\ 5 & 9 & 8 & 11 & 10 \end{bmatrix}.$$

3. (From Chapter 3, "Discrete Location Theory", Ed., Mirchandani and Francis.) Let P denote the set of nine integral points in the square $0 \le x \le 2,\ 0 \le y \le 2$. For any point $i \in P$, let x_i and y_i be its coordinates. Consider the instance of the UFL problem defined by $m = n = 9$, $f_j = 2$ for all $j \in P$, and $c_{ij} = -|x_i - x_j| - |y_i - y_j|$ for all $i,\ j \in P$.

   (a) Use the greedy heuristic to find the values Z^G and W^G.

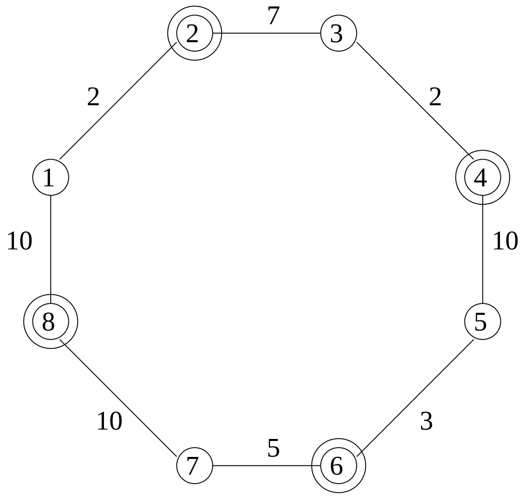
   (b) Apply the the dual descent procedure and find the value $W^{DD}(u)$. Is the set $K(u)$ an optimal set of open facilities for this problem instance?

4. Apply the Aardal-Shmoys-Tardos algorithm (based on the Filtering & Rounding method) to the following "minimize" UFL problem:

$$V = \{1, 2, 3, 4, 5, 6, 7, 8\}, \quad I = \{1, 3, 5, 7\}, \quad J = \{2, 4, 6, 8\}.$$

   The fixed costs f_j for the sites $j \in J$ are

$$f_2 = 400, \quad f_4 = 800, \quad f_6 = 400, \quad f_8 = 200.$$



$$[c_{vw}] = \begin{bmatrix} 0 & 2 & 9 & 11 & 21 & 24 & 20 & 10 \\ 2 & 0 & 7 & 9 & 19 & 22 & 22 & 12 \\ 9 & 7 & 0 & 2 & 12 & 15 & 20 & 19 \\ 11 & 9 & 2 & 0 & 10 & 13 & 18 & 21 \\ 21 & 19 & 12 & 10 & 0 & 3 & 8 & 18 \\ 24 & 22 & 15 & 13 & 3 & 0 & 5 & 15 \\ 20 & 22 & 20 & 18 & 8 & 5 & 0 & 10 \\ 10 & 12 & 19 & 21 & 18 & 15 & 10 & 0 \end{bmatrix}$$

Figure 4.5: Graph $G = (V, E)$ and cost matrix $[c_{vw}]$ for Exercise 4.

The service costs are given by the matrix $[c_{vw}]$ $(\forall v, w \in V \times V)$ in Figure 4.5. Note that the matrix $[c_{vw}]$ is symmetric and satisfies the triangle inequality.

Start with the fractional solution $\widehat{y}$ given by

$$\widehat{y}_2 = 1/4, \quad \widehat{y}_4 = 1/4, \quad \widehat{y}_6 = 1/4, \quad \widehat{y}_8 = 1/4.$$

Take the parameter $\epsilon = 1/2$.

5. The aim here is to consider a generalization of the "minimize" UFL problem and to develop a 4-approximation algorithm.

In the "minimize" UFL problem, suppose that each client $i \in I$ has a nonnegative real-valued demand h_i (different clients may have different demands). Consider the modified objective function

$$\text{minimize} \quad \sum_{i \in I} \sum_{j \in J} h_i c_{ij} x_{ij} + \sum_{j \in J} f_j y_j.$$

Modify the Aardal-Shmoys-Tardos algorithm so that it finds a feasible integral solution $\widetilde{x}, \widetilde{y}$ whose objective value is $\leq 4 z_{LP}$, where z_{LP} is the objective value of the LP relaxation using the above objective function.

Bibliography

[1] G. Cornuejols, G. L. Nemhauser and L. A. Wolsey, *The uncapacitated facility location problem.* in P. Mirchandani and R. Francis, editors, *Discrete Location Theory*, John Wiley and Sons Inc., New York, 1990, pp. 119–171.

[2] R. C. Larson and A. R. Odoni, *Urban Operations Research.* Prentice-Hall, Englewood Cliffs, N. J., 1981.

[3] J.-H. Lin and J. S. Vitter, *ϵ-approximations with minimum packing constraint violation*, Proc. 24th Annual ACM Symp. on Theory of Computing (1992), pp. 771–782.

[4] J.-H. Lin and J. S. Vitter, *Approximation algorithms for geometric location problems*, Inf. Proc. Lett. **44** (1992), pp. 245–249.

[5] P. B. Mirchandani and R. L. Francis, *Discrete Location Theory.* John Wiley & Sons, New York, 1990.

[6] G. L. Nemhauser and L. A. Wolsey, *Integer and Combinatorial Optimization*, John Wiley and Sons Inc., New York, 1988.

[7] P. Raghavan, *Probabilistic construction of deterministic algorithms: approximating packing integer programs*, J. Comp. Sys. Sci. **37** (1988), pp. 130–143.

[8] P. Raghavan and C. D. Thompson, *Randomized rounding: a technique for provably good algorithms and algorithmic proofs*, Combinatorica **7** (1987), pp. 365–374.

[9] D. B. Shmoys, E. Tardos and K. Aardal, *Approximation algorithms for facility location problems.* Proc. 29th ACM Sympos. on Theory of Computing (1997), pp. 265–274.