Faculty of Mathematical Sciences

University of Twente

University for Technical and Social Sciences

P.O. Box 217
7500 AE Enschede
The Netherlands
Phone: +31-53-4893400
Fax: +31-53-4893114
Email: memo@math.utwente.nl

Memorandum No. 1540

An approximation algorithm for the 2-level uncapacitated facility location problem

A.F. Bumb

September 2000

ISSN 0169-2690

An approximation algorithm for the 2-level uncapacitated facility location problem

Adriana Bumb
Faculty of Mathematical Sciences
University of Twente

Abstract

We present an approximation algorithm for the maximization version of the two level uncapacitated facility location problem achieving a performance guarantee of 0.47. The main idea is to reduce the problem to a special case of MAX SAT, for which an approximation algorithm based on randomized rounding is presented.

Key words: approximation algorithms, randomized algorithms, satisfiability problem, facility location
AMS Subject Classifications: 68W25, 68W20, 90B80

1 Introduction

The two level uncapacitated facility location problem (two level UFLP), in the maximization version, can be described as follows. There are two types of potential facility locations : the hub facilities, denoted by F and the transit facilities, denoted by E. Building and opening the facility $i \in F$ has an associated nonnegative cost f_i and $j \in E$ a nonnegative cost e_j. There is also a set of clients, D, who should be assigned to open pairs of facilities from $F \times E$. If a client $k \in D$ is assigned to the pair (i, j), a profit c_{ijk} is incurred.The problem is to decide simultaneously which facilities from F and which from E to open (at least one from each set) and how to assign the clients to the open facilities, such that the total profit is maximized.

Formally, the problem can be stated as

$$\max_{S_1 \times S_2 \subseteq F \times E} C(S_1, S_2) = \sum_{k \in D} \max_{(i,j) \in S_1 \times S_2} c_{ijk} - \sum_{i \in S_1} f_i - \sum_{j \in S_2} e_j.$$

Denote by $C_R = \sum_{k \in D} \min_{(i,j) \in F \times E} c_{ijk} - \sum_{i \in F} f_i - \sum_{j \in E} e_j$ and by C^* the optimal value of the problem.

Throughout this paper, a $\rho-$*approximation algorithm* is a polynomial time algorithm that always finds a feasible solution with objective function value at least ρ times the optimum. The value ρ is called the *performance guarantee* of the algorithm.

If the set E is empty, one obtains the one level uncapacitated facility location problem. Cornuejols, Fisher and Nemhauser [4] proved that for this NP-hard problem a simple greedy algorithm finds a solution S such that

$$C(S) - C_R \geq (1 - \frac{1}{e})(C^* - C_R).$$

Recently, Ageev and Sviridenko, [2], improved the performance guarantee to 0.828. Their algorithm has two steps: in the first one they reduce the one level uncapacitated facility location problem to a special case of the maximum satisfiability problem (MAX SAT*) and in the second they find an 0.828$-$approximation algorithm for the MAX SAT* applying a generalization of the randomized algorithm proposed in [6] for the general MAX SAT problem.

Being a generalization of the one level uncapacitated facility location problem, the two level UFLP is NP-hard as well .Only a few algorithms for the two level UFLP have been developed (see Aardal et al.[1] for a survey).The techniques which have been used are branch-and-bound, Lagrangean relaxation, cutting planes.

In this paper we describe a polynomial time approximation algorithm for solving the two level UFLP problem based on the technique of randomized rounding. We prove that the algorithm delivers a solution $S_1 \times S_2 \subseteq F \times E$ such that

$$C(S_1, S_2) - C_R \geq 0.47(C^* - C_R).$$

The algorithm is based on the observation that the two level UFLP in the "shifted form",i.e.

$$\max_{S_1 \times S_2 \subseteq F \times E} (C(S_1, S_2) - C_R)$$

admits a two-way approximation preserving reduction to a special case of MAX SAT.

2 Approximation preserving reductions between two level UFLP and MAX SAT SP

In this section we present a way to reduce the two level UFLP problem to a special case of MAX SAT,called MAX SAT SP and the reverse reduction, such that the relative errors of the corresponding feasible solutions are preserved. The reductions follow the same line as the reduction of the one level uncapacitated facility location problem to MAX SAT, proposed by Ageev&Sviridenko in [2] .

We introduce a special subclass MAX SAT SP of the wellknown MAX SAT problem.

In MAX SAT SP, there are two sets of boolean variables $\{y_{1,...}y_m\}$ and $\{z_{1,...,}z_n\}$ and L sets $I_{1,...,}I_L$,included in $\{1, ..., m\} \times \{1, ..., n\}$ .For each y_i ,

$i = \overline{1,m}$ there is a clause of the form $\overline{y_i}$,of weight u_i.Similarly, for each z_j , $j = \overline{1,n}$ there is a clause of the form $\overline{z_j}$,of weight v_j.Finally, for each I_l , $l = \overline{1,L}$, there is a clause of the form $\bigvee_{(i,j)\in I_l} (y_i \wedge z_j)$ of weight w_l.

Given $m, n,$ the sets $I_{1,...,}I_L$, the weights w_l, for $l = \overline{1,L}$, u_i, for $i = \overline{1,m}$ and v_j for $j = \overline{1,n}$,one has to find an assignment of true values to $y_{1,...}y_m$ and to $z_{1,...,}z_n$ maximizing the total weight of satisfied clauses.

In this paper we will identify the truth and false values with 1 and 0 respectively. Then,

MAX SAT SP can be formulated as follows

$$\max_{\substack{y\in\{0,1\}^m,\\ z\in\{0,1\}^n}} \sum_{l=1}^{L} w_l(\bigvee_{(i,j)\in I_l} (y_i \wedge z_j)) + \sum_{i=1}^{m} u_i\overline{y_i} + \sum_{j=1}^{n} v_j\overline{z_j}, \tag{1}$$

Next we will present the reductions between the two level UFLP problem and the MAX SAT SP.

Consider an instance of the two level UFLP problem.

For each $k \in D$, suppose that

$$c_{i_1(k)j_1(k)k} \geq c_{i_2(k)j_2(k)k} \geq ...c_{i_p(k)j_p(k)k},$$

where $p = |F| \times |E|$.

For every $s \in \{1, ...p\}$ define the sets I_{sk} as being the set of the s most profitable pairs for k

$$I_{sk} = \{(i_1(k), j_1(k)), ..., (i_s(k), j_s(k))\}$$

With each set defined above we associate a number w_{sk}

$$w_{sk} = c_{i_s(k)j_s(k)k} - c_{i_{s+1}(k)j_{s+1}(k)k}, \text{ for } s \leq p-1$$

and

$$w_{pk} = 0.$$

For the sets $S_1 \subseteq F$ and $S_2 \subseteq E$ let $y = \{y_{1,...}y_m\}$ and $z = \{z_{1,...,}z_n\}$ be the incidence vectors of S_1, respectively S_2.The objective function of the two level UFLP problem can be rewritten as

$$\begin{aligned}
\sum_{k\in D} \max_{(i,j)\in S_1\times S_2} c_{ijk} - \sum_{i\in S_1} f_i - \sum_{j\in S_2} e_j = \sum_{k\in D} \left[\sum_{s=1}^{p} w_{sk} \left(\bigvee_{(i,j)\in I_{sk}} (y_i \wedge z_j) \right) + \min_{(i,j)\in F\times E} c_{ijk} \right] \\
+ \sum_{i\in F} f_i\overline{y_i} + \sum_{j\in E} e_j\overline{z_j} - \sum_{i\in F} f_i - \sum_{j\in E} e_j \\
= \sum_{k\in D} \left[\sum_{s=1}^{p} w_{sk} \left(\bigvee_{(i,j)\in I_{sk}} (y_i \wedge z_j) \right) \right] + \sum_{i\in F} f_i\overline{y_i} + \sum_{j\in E} e_j\overline{z_j} \\
+ \sum_{k\in D} \min_{(i,j)\in F\times E} c_{ijk} - \sum_{i\in F} f_i - \sum_{j\in E} e_j.
\end{aligned}$$

Hence,

$$C(S_1,S_2)-C_R=\sum_{k\in D}\left[\sum_{s=1}^{p} w_{sk}\left(\bigvee_{(i,j)\in I_{sk}}(y_i\wedge z_j)\right)\right]+\sum_{i\in F} f_i\overline{y_i}+\sum_{j\in E} e_j\overline{z_j},$$

which implies that the two level UFLP problem reduces to MAX SAT SP in the form (1).

The backward reduction is as follows. Let $(m,n,L,I_{1,...,}I_L,$ $w_1,...,w_L,u_1...u_m,v_{1,...}v_n)$ be an instance of MAX SAT SP. If we apply to the instance of the two level UFLP with $F=\{1,...,m\}, E=\{1,...,n\}, D=\{1,...,L\}, f_i=u_i$ for $i=\overline{1,m}, e_j=v_j$ for $j=\overline{1,n}$ and

$$c_{ijl}=\begin{cases} w_l \text{ if } (i,j)\in I_l \\ 0 \text{ otherwise}\end{cases} \text{, for each } l\in L \text{ and } (i,j)\in F\times E$$

the reduction described previously, we will arrive to the original instance of MAX SAT SP.

Thus, the described reductions between two level UFLP and MAX SAT SP preserve relative errors of the corresponding feasible solutions.

3 An 0.47 approximation algorithm for MAX SAT SP

In this section we will present a 0.47 approximation algorithm for MAX SAT SP, based on independent randomized rounding.

The MAX SAT SP problem can be formulated as the following integer program:

$$\max\sum_{l=1}^{L} w_l t_l+\sum_{i=1}^{m} u_i(1-y_i)+\sum_{j=1}^{n} v_j(1-z_j)$$

$$s.t.\ t_l\le\sum_{(i,j)\in I_l} x_{ij}\ \text{, for each } l=1,...L \quad (2)$$

$$x_{ij}\le y_i \text{ , for each } i=1,...,m \text{ , } j=1,...,n \quad (3)$$

$$x_{ij}\le z_j \text{ , for each } i=1,...,m \text{ , } j=1,...,n \quad (4)$$

$$x_{ij}\in\{0,1\}\text{, for each } i=1,...,m \text{ , } j=1,...,n \quad (6)$$

$$y_i\in\{0,1\}\text{, for each } i=1,...,m$$

$$z_j\in\{0,1\}\text{, for each } j=1,...,n$$

$$t_l\in\{0,1\}\text{, for each } l\in L$$

For each $i=1,...,m$, $j=1,...,n$ the variable x_{ij} substitutes $y_i\wedge z_j$, i.e. takes value 1 if both y_i and z_j are 1 and 0 otherwise. For each $l=1,...L$, the variable t_l substitutes $\bigvee_{(i,j)\in I_l}(y_i\wedge z_j)$, so it takes value 1 if there is a pair

$(i, j) \in I_l$ such that both y_i and z_j take value 1;if there is no such pair, t_l will be 0.

The Algorithm

We will consider the LP relaxation of the integer programming formulation of MAX SAT SP with all variables taking values in the interval $[0, 1]$.

Let $(\widetilde{x}, \widetilde{y}, \widetilde{z}, \widetilde{t})$ be an optimal solution of the LP relaxation, and let $\widetilde{C_{LP}}$ be its optimal value, i.e.

$$\widetilde{C_{LP}} = \sum_{l=1}^{L} w_l \widetilde{t_l} + \sum_{i=1}^{m} u_i(1 - \widetilde{y_i}) + \sum_{j=1}^{n} v_j(1 - \widetilde{z_j}).$$

Clearly, $\widetilde{C_{LP}}$ is an upperbound of the optimal value of MAX SAT SP.

Consider a $\lambda \in [0, 1]$. Next the algorithm independently sets each y_i to 1 with probability $p_i = (1 - \lambda) + \lambda \widetilde{y_i}$ and to 0 with probability $1 - p_i = \lambda(1 - \widetilde{y_i})$. Similarly, each z_j will take value 1 with probability $q_j = (1 - \lambda) + \lambda \widetilde{z_j}$ and value 0 with probability $1 - q_j = \lambda(1 - \widetilde{z_j})$. For each $(i, j) \in \{1, ..., m\} \times \{1, ..., n\}$, set x_{ij} to 1 if both y_i and z_j were set to 1. For each $l \in \{1, ..., L\}$, set t_l to 1 if there is a pair $(i, j) \in I_l$ such that x_{ij} was previously set to 1.

In this way we obtain for every value of λ in $[0, 1]$a feasible solution of the integer program.

Denote with $C(\lambda)$ the expected value of the algorithm. To prove the performance guarantee of the algorithm we will compare $C(\lambda)$ with $\widetilde{C_{LP}}$.

Theorem 1 *The expected value of the algorithm is at least* $0.47\ \widetilde{C_{LP}}$.

Proof

From the linearity of the expectation it follows

$$\begin{aligned} C(\lambda) &= \sum_{l=1}^{L} w_l Prob(t_l = 1) + \sum_{i=1}^{m} u_i(1 - Prob(y_i = 1)) + \sum_{j=1}^{n} v_j(1 - Prob(z_j = 1)) \\ &= \sum_{l=1}^{L} w_l Prob(t_l = 1) + \lambda \sum_{i=1}^{m} u_i(1 - y_i) + \lambda \sum_{j=1}^{n} v_j(1 - z_j). \end{aligned}$$

To calculate the probabilities that t_l take value 1 we will study more cases, dependent on the structure of the sets I_l.The main idea is that the events of choosing the value 0 or 1 for y_i's and z_j's are independent.

Case 1. $I_l = \{(i, j)\}$

$$\begin{aligned} Prob(t_l = 1) &= Prob((y_i = 1) \wedge (z_j = 1)) = Prob(y_i = 1) Prob(z_j = 1) \\ &= [(1 - \lambda) + \lambda \widetilde{y_i}][(1 - \lambda) + \lambda \widetilde{z_j}] \end{aligned}$$

Using the inequality

$$a + b \geq 2\sqrt{ab}$$

we can obtain the following lowerbound

$$Prob(t_l = 1) \geq 4(1-\lambda)\lambda\sqrt{\widetilde{y_i}\widetilde{z_j}} \geq 4(1-\lambda)\lambda\widetilde{t_l}$$

Case 2.$I_l = \{(i, j_1), ..., (i, j_s)\}, s \geq 2$

$$\begin{aligned} Prob(t_l = 1) &= Prob(y_i = 1)Prob((z_{j_1} = 1) \vee ... \vee (z_{j_s} = 1)) \\ &= Prob(y_i = 1)(1 - Prob((z_{j_1} = 0) \wedge ... \wedge (z_{j_s} = 0)) \\ &= Prob(y_i = 1)(1 - \prod_{q=1}^{s} Prob(z_{j_q} = 0)) \\ &= [(1-\lambda) + \lambda\widetilde{y_i}][1 - \lambda^s \prod_{q=1}^{s}(1 - \widetilde{z_{j_q}})] \end{aligned}$$

The arithmetic/geometric mean inequality, applied to $1 - \widetilde{z_{j_q}}, \overline{q = 1, s}$ gives

$$\prod_{q=1}^{s}(1 - \widetilde{z_{j_q}}) \leq \left(1 - \frac{\sum_{q=1}^{s} \widetilde{z_{j_q}}}{s}\right)^s.$$

Hence, we obtain the following lowerbound for $Prob(t_l = 1)$:

$$Prob(t_l = 1) \geq [(1-\lambda) + \lambda\widetilde{y_i}][1 - \lambda^s \left(1 - \frac{\sum_{q=1}^{s} \widetilde{z_{j_q}}}{s}\right)^s].$$

The function $f : [0,1] \to \mathcal{R}$ defined by

$$f(x) = 1 - a\left(1 - \frac{x}{s}\right)^s, \text{ where } a \in [0,1],$$

is concave. Observing that any $x \in [0,1]$ can be written as a convex combination between 0 and 1, the concavity of f implies

$$f(x) = f(1 * x + 0 * (1-x)) \geq x * f(1) + (1-x) * f(0).$$

From

$$\begin{aligned} f(0) &= 1 - a \geq 0, \\ f(1) &= 1 - a\left(1 - \frac{1}{s}\right)^s, \end{aligned}$$

it follows that

$$f(x) \geq [1 - a\left(1 - \frac{1}{s}\right)^s]x.$$

Substituting in this inequality $a = \lambda^s$, $x = \sum_{q=1}^{s} \widetilde{z_{j_q}}$ and taking into account that by (2) and (4)
$\sum_{q=1}^{s} \widetilde{z_{j_q}} \geq \widetilde{t_l}$,we obtain

$$1 - \lambda^s \left(1 - \frac{\sum_{q=1}^{s} \widetilde{z_{j_q}}}{s}\right)^s \geq [1 - \lambda^s \left(1 - \frac{1}{s}\right)^s] \sum_{q=1}^{s} \widetilde{z_{j_q}} \geq [1 - \lambda^s \left(1 - \frac{1}{s}\right)^s]\widetilde{t_l}.$$

Thus, for this case the lowerbound for the probability of t_l being 1 is:

$$Prob(t_l = 1) \geq [(1-\lambda) + \lambda\widetilde{y_i}][1 - \lambda^s \left(1 - \frac{1}{s}\right)^s]\widetilde{t_l} \geq (1-\lambda)[1 - \lambda^s \left(1 - \frac{1}{s}\right)^s]\widetilde{t_l}.$$

Case 3. $I_l \supseteq \{(i_1, j_1), (i_2, j_2)\}$ with $i_1 \neq i_2$ and $j_1 \neq j_2$.
In this case, the event that the pair (i_1, j_1) is open is independent of the event that the pair (i_2, j_2) is open and consequently,

$$\begin{aligned} Prob(t_l = 1) &\geq Prob\left[(y_{i_1} = 1 \wedge z_{j_1} = 1) \vee (y_{i_2} = 1 \wedge z_{j_2} = 1)\right] \\ &= Prob\left(y_{i_1} = 1 \wedge z_{j_1} = 1\right) + Prob\left(y_{i_2} = 1 \wedge z_{j_2} = 1\right) \\ &- Prob\left(y_{i_1} = 1 \wedge z_{j_1} = 1\right) Prob\left(y_{i_2} = 1 \wedge z_{j_2} = 1\right) \\ &= p_{i_1}q_{j_1} + p_{i_2}q_{j_2} - p_{i_1}q_{j_1}p_{i_2}q_{j_2} \\ &\geq 2\sqrt{p_{i_1}q_{j_1}p_{i_2}q_{j_2}} - p_{i_1}q_{j_1}p_{i_2}q_{j_2}. \end{aligned}$$

By the definition of p_i and q_j , $p_i \geq 1 - \lambda$ and $q_j \geq 1 - \lambda$ for each i and j.Hence,

$$p_{i_1}q_{j_1}p_{i_2}q_{j_2} \geq (1-\lambda)^4.$$

The function $f : \mathcal{R}_+ \to \mathcal{R}_+$ defined by $f(x) = 2\sqrt{x} - x$ is increasing on $[0, 1]$, which together with the inequality above implies that

$$Prob(t_l = 1) \geq 2\sqrt{(1-\lambda)^4} - (1-\lambda)^4 \geq [2(1-\lambda)^2 - (1-\lambda)^4]\widetilde{t_l}.$$

¿From the cases $(1) - (3)$ it follows that for a $\lambda \in [0, 1]$,

$$C(\lambda) \geq \min\{\lambda, 4\lambda(1-\lambda), 2(1-\lambda)^2 - (1-\lambda)^4, \min_{s \geq 1}(1-\lambda)[1 - \lambda^s \left(1 - \frac{1}{s}\right)^s]\}\widetilde{C_{LP}}.$$

Hence, the performance guarantee of the algorithm is equal with

$$\rho = \max_{\lambda \in [0,1]} \min\{\lambda, 4\lambda(1-\lambda), 2(1-\lambda)^2 - (1-\lambda)^4, \min_{s\geq 1}(1-\lambda)[1-\lambda^s\left(1-\frac{1}{s}\right)^s]\}$$

Using the fact that $\left(1-\frac{1}{s}\right)^s \leq e^{-1}$, for every $s \geq 1$, we obtain that $\rho = 0.47$, for $\lambda = 0.47$.

■

Remark

The randomized algorithm presented above can be derandomized using the method of conditional expectations [3]. The result is a deterministic algorithm which finds in polynomial time a solution with a value at least 0.47 the optimum.

References

[1] K. Aardal, M. Labbe, J. Leung , M. Queyranne. On the two-level uncapacitated facility location problem. *INFORMS Journal on Computing, 8, pages 289-301, 1996*

[2] A. Ageev, M. Sviridenko. An 0.828-approximation algorithm for uncapacitated facility location problem. Discrete Applied Mathematics, 93, pages 289-296, 1999

[3] N. Alon, J.H.Spencer and P.Erdös. The probabilistic method. John Wiley and Sons, New York, 1992

[4] G. Cornuejols, M.L.Fisher and G.L. Nemhauser. Location of bank accounts to optimize float: an analytic study of exact and approximate algorithms, *Management Science*, 23, *pages 789-810, 1977*

[5] G. Cornuejols, G.L. Nemhauser and L.A. Wolsey. The uncapacitated facility location problem. In P. Mirchandani and R. Francis, editors, *Discrete Location Theory*, pages 119-171. John Wiley and Sons, New York, 1990

[6] M.X.Goemans,D.P.Williamson. A new 3/4-approximation algorithm for the Maximum Satisfiability Problem. *SIAM Journal on Discrete Mathematics, 7, pages 656-666, 1994*