Working paper # 25-80-81

Location Problems, Set Covering Problems and the Greedy Algorithm

by

Gerard Cornuejols

and

Jean-Michel Thizy

Carnegie-Mellon University

This work was supported by NSF Grant ENG 7902506.

1. INTRODUCTION

Cornuejols, Fisher and Nemhauser [2] proved that a greedy solution to the simple plant location problem never deviates from the optimal solution by more than $\frac{1}{e}$ (in relative error), where e is the base of natural logarithms. For the set covering problem Lovasz [5], Johnson [6] and Chvatal [1] gave a logarithmic bound for the ratio of a greedy solution to the optimal solution. Recently Fisher [4] pointed to the close relationship between these two results.

In this paper we strengthen the bound of [2] in two different ways and we show how these new results relate to the tight bounds proved by Lovasz, Johnson and Chvatal.

The weighted plant location problem: a simple plant location problem with one budget constraint.

The demands at n locations must be satisfied at minimum cost by investing in some facilities chosen among m possible sites within a given budget.

The problem can be formulated as:

$$z = \max \quad \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$$

$$(1) \qquad \sum_{j=1}^{n} x_{ij} = 1 \qquad i = 1, ---, m$$

$$(2) \qquad \sum_{j=1}^{n} w_j y_j \le k$$

$$(3) \qquad 0 \le x_{ij} \le y_j \le 1 \qquad i = 1, ---, m; \quad j = 1, ---, n$$

$$(4) \qquad y_j \text{ integral} \qquad j = 1, ---, n \; ^*$$

---

*Note from J.-M. Thizy: j = 1, ---, n omitted from the original Working paper.

<u>where</u>: y_j = 1 if facility is opened at location j

0 otherwise

x_{ij} = 1 if facility (j) satisfies demand (i).

0 otherwise

w_j is the investment cost of facility (j)

c_{ij} is the (negative) cost of satisfying demand (i) from location (j).

<u>Note</u>: Since the sign of c_{ij} is not specified, the problem can have a minimizing or maximizing objective.

The weighted plant location problem subsumes two well-known problems:

The simple K-plant location problem ($w_j = 1$, $j = 1, ---, n$) (5)

the 0-1 knapsack problem ($m = n$, $c_{ij} = 0$ for $j \neq i$, $i = 1, ---, m$, $j = 1, ---, n$)

In the sequel, c_{ij} will be called the costs, w_j the weights, and K the weight limit of the problem. We also restrict w_j to be nonnegative (if w_j was negative, the corresponding location would always be opened, and the demands thus satisfied could be omitted in the formulation. The restricted problem with these locations and corresponding clients omitted is equivalent). Note also that if $w_j > K$ the corresponding location can never be chosen in a feasible solution. So we assume from now on that $0 \leq w_j \leq K$ for every j. Similar remarks can be made for the knapsack problem.

2. The weighted plant location problem: worst-case analysis of a greedy algorithm.

A Lagrangian relaxation:

Let $u = (u_1, ---, u_m)$ be multipliers for the constraints (1), and define

$$L(x, y, u) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} - \sum_{i=1}^{m} u_i \left(\sum_{i=1}^{m} x_{ij} - 1 \right)$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{m} (c_{ij} - u_i) x_{ij} + \sum_{i=1}^{m} u_i$$

The Lagrangian problem is:

$$P(u) \qquad z_D(u) = \max_{x,\,y} L(x, y, u)$$

subject to (2) - (3) - (4)

Define $z_D = \min_u z_D(u)$. It is well-known that $z \leq z_D$.

P(u) is a 0-1 knapsack problem, for which an analytic upper bound is available.

Define $\rho_j(u) = \sum_{i=1}^{m} \max(0, c_{ij} - u_i)$

and order the ratios $\dfrac{\rho_j(u)}{w_j}$ in non increasing order:

$$\frac{\rho_{p(u,k)}}{w_{p(u,k)}} \geq \frac{\rho_{p(u,h)}}{w_{p(u,h)}} \quad \text{if } k < h$$

then

$$z \leq z_D \leq z_D(u) \leq \sum_{i=1}^{m} u_i + \sum_{h=1}^{r(u)-1} \rho_{p(u,h)} + \frac{\rho_{r(u)}}{w_{r(u)}} \left[K - \sum_{h=1}^{r(u)-1} w_{p(u,h)} \right] \tag{6}$$

where r(u) is the minimum value satisfying:

$$\sum_{h=1}^{r(u)-1} w_{p(u,h)} \leq K \leq \sum_{h=1}^{r(u)} w_{p(u,h)}$$

The greedy heuristic

The greedy heuristic chooses facilities one at a time. At each step the chosen facility is one that produces the largest improvement in the objective value per unit of investment.

To formalize this statement, the following algorithm is given:

Greedy location algorithm

k counts the number of locations chosen (recorded in J*)

Step 1 Let $k = 1$, $J* = \emptyset$ and $u_i^1 = \min\limits_{j=1,---,n} c_{ij}$ for $i = 1, ---, m$

Step 2 Let $\rho_j(u^k) = \sum\limits_{i=1}^{m} \max\left(0, c_{ij} - u_i^k\right)$, $j \notin J*$

Find $j_k \notin J*$ such that $\dfrac{\rho_{j_k}(u^k)}{w_{j_k}} = \max\limits_{j \notin J*} \dfrac{\rho_j(u^k)}{w_j}$.

If $\sum\limits_{h=1}^{k} w_{j_h} > K$ or $\rho_{j_k}(u^k) \leq 0$ and $|J*| \geq 1$ go to Step 4.

Otherwise set $J* = J* \cup \{j_k\}$ and go to Step 3.

Step 3 Set $k = k + 1$. For $i = 1, ---, m$, set $u_i^k = \max\limits_{j \varepsilon J*} c_{ij}$

$= u_i^{k-1} + \max\left(0, c_{ij_{k-1}} - u_i^{k-1}\right)$

Go to Step 2.

Step 4 Stop; set $H = k-1$, the greedy solution is given by $y_j = 1$, $j \varepsilon J*$ and $y_j = 0$ otherwise, and the value of the greedy solution is

$$z_G = \sum_{i=1}^{m} u_i^1 + \sum_{h=1}^{H} \rho_{j_h} .$$

In this equation and in the sequel of the paper ρ_{j_h} denotes $\rho_{j_h}(u^h)$.

Let $z_R = \sum_{i=1}^{M} u_i^1$

<u>Theorem 1</u>: for all weighted plant location problems (1) - (4), a heuristic solution with value z_G and number of steps H satisfies:

$$(7) \qquad \frac{z_D - z_G}{z_d - z_R} \leq \prod_{h=1}^{H} \left(1 - \frac{w_{j_h}}{k}\right)$$

Proof: We first show that:

$$(8) \qquad \sum_{i=1}^{m} u^h = \sum_{i=1}^{m} u_i^1 + \sum_{k=1}^{h-1} \rho_{j_k}$$

this is because

for i=1, ---, m $\qquad u_i^k = u_i^{k-1} + \max\left(0,\ c_{ij_{k-1}} - u_i^{k-1}\right)$

therefore $\qquad \sum_{i=1}^{m} u_i^k = \sum_{i=1}^{m} u_i^{k-1} + \sum_{i=1}^{m} \max\left(0,\ c_{ij_{k-1}} - u_i^{k-1}\right)$

the last term of the second member is $\rho_{j_{k-1}}\left(u^{k-1}\right) = \rho_{j_{k-1}}$

therefore

$$(9)_k \qquad \sum_{i=1}^{m} u_i^k = \sum_{i=1}^{m} u_i^{k-1} + \rho_{j_{k-1}}$$

and adding each member of $(9)_k$ for k=2, ---, h, and cancelling the terms

$$\sum_{i=1}^{m} u_i^k = \sum_{i=1}^{m} u_i^1 + \rho_{j_k} + \rho_{j_{k-1}} + --- + \rho_{j_l}$$

Show that:

$$(10) \qquad \frac{\rho_{j_h}}{w_{j_h}} \geq \frac{\rho_j(u^h)}{w_j} \qquad \begin{array}{l} j = 1, ---, n \\ h = 1, ---, H \end{array}$$

(11) by definition $\rho_j(u) \geq 0 \quad j = 1, ---, n$

also, for $j \varepsilon J*(h)$ (J*(h) is J* at step h of greedy heuristic)

$$\rho_j(u^h) = \sum_{i=1}^{m} \max (0, c_{ij} - u_i^h)$$

but

$$u_i^h = \max_{j \varepsilon J*(h)} c_{ij}, \text{ therefore } c_{ij} - u_i^h \leq \underset{i=1}{0}, \; j \varepsilon \underset{---,\, m}{J*(h)}$$

(12) Therefore $\rho_j(u^h) = 0, \; j \varepsilon J*(h)$

recall:

(13) $$\frac{\rho_{j_h}}{w_{j_h}} = \max_{j \varepsilon J*(h)} \frac{\rho_j(u^h)}{w_j}$$

and (11, (12), (13) imply (10).

Recall (6): for $h = 1, ---, H \leq n$

$$z_D \leq z_D(u^h) \leq \sum_{i=1}^{m} u^h + \sum_{j=1}^{r(u^h)} \rho_{p(u^h,j)} + \left[K - \sum_{j=1}^{r(u^h)} w_{p(u^h,j)}\right] \frac{\rho_{r(u^h)}}{w_{r(u^h)}}$$

from (8) and (10)

$$z_D \leq \sum_{i=1}^{m} u_i^1 + \sum_{k=1}^{h-1} \rho_{j_k} + \sum_{k=1}^{r(u^h)} w_{j_h} \frac{\rho_{j_k}}{w_{j_k}} + \left(K - \sum_{k=1}^{r(u^h)} w_{j_k}\right) \frac{\rho_{j_h}}{w_{j_h}}$$

$$\leq z_R + \sum_{k=1}^{h-1} \rho_{j_k} + \frac{K\rho_{j_h}}{w_{j_h}}$$

In the sequel of the paper ρ_h, w_h will be used instead of ρ_{j_h}, w_{j_h} for notational simplicity. For a given number of heuristic iterations H, the value of (ρ_j, j=1, ---, n) that maximizes $\frac{z_D - z_G}{z_D - z_R}$ is found by solving the program:

$$v_H = \max \frac{z_D - z_G}{z_D - z_R} = 1 - \frac{1}{z_D - z_R} \sum_{k=1}^{H} \rho_k$$

(14)$_h$ $$z_D - z_R \le \sum_{k=1}^{h-1} \rho_k + \frac{K\rho_h}{w_h} \qquad h = 1, ---, H$$
$$k = 1, ---, H$$

$$\rho_k \ge 0$$

where ρ_k, $k = 1$, ---, H are the variables. Let $w'_h = \frac{w_h}{K}$.

The following solution satisfies all constraints (14)h as equalities

$$\rho_h = (z_D - z_R)\, w'_h \prod_{k=1}^{h-1} (1 - w'_k) \qquad h = 1, ---, H$$

a dual solution satisfies the non-trivial constraints as equalities,

$$t_h = \frac{w'_h}{z_D - z_R} \prod_{k=h-1}^{H} (1 - w'_k) \qquad h = 1, ---, H$$

(with the convention $\prod_{j=n}^{n-1} (1 - w'_j) = 1$ for any n)

these two solutions, satisfying the complementary slackness conditions, are therefore optimal.

Note that $(14)_H$ is:

$$Z_D \le \sum_{h=1}^{H-1} \rho_h + \frac{\rho_H}{w'_H} = \sum_{h=1}^{H} \rho_h + \rho_H \left(\frac{1}{w'_H} - 1\right)$$

therefore the optimal value of the program is

$$v_H = 1 - \frac{1}{z_D - z_R} \left[z_D - \rho_H \left(\frac{1}{w'_H} - 1 \right) \right]$$

$$= \left(\frac{1}{w'_H} - 1 \right) w'_H \prod_{h=1}^{H-1} (1 - w'_h)$$

$$= \prod_{h=1}^{H} (1 - w'_h)$$

<u>Corollary 1:</u> For the K-plant location problem, the greedy heuristic satisfies

$$\frac{Z_D - Z_G}{Z_D - Z_R} \leq \left(\frac{K-1}{K}\right)^K \left(< \frac{1}{e}\right)$$

Thus theorem 1 generalizes a theorem of Cornuejols, Fisher and Nemhauser [2].

Recall that for notational simplicity we will use ρ_h and w_h instead of $\rho_{j_h}\left[= \rho_{j_h}\left(u^H\right)\right]$ and w_{j_h} in the remainder of the paper.

3. The greedy covering and greedy location heuristics.

Cornuejols, Nemhauser and Wolsey [3] have shown a method of reducing a simple plant location problem with general costs c_{ij} to another simple plant location problem with zero-one costs. The constructive proof also holds for the weighted case.

When the matrix C is a 0, 1 matrix, the relationship between the set covering problem and the weighted plant location problem can be captured by the following summary descriptions.

Set covering: Cover all the rows of C by a set of columns with minimum weight.

Weighted plant location problem: Cover a maximum number of rows by a set of columns with specified weight limit.

More precisely, let $C = (c_{ij})$ be a m x n matrix where $c_{ij} = 0$ or 1,

$w = (w_j)$ be a row vector with h components,

$e = (1, 1, 1, 1, ---, 1)^T$ be a column vector with m ones,

$x = (x_j)$ be a column vector with n 0-1 components.

The set covering problem can be written as:

$$(15) \qquad s = \min\ w\,x$$
$$C\,x \geq e$$
$$x \text{ binary}$$

A compact reformulation of the weighted plant location problem (1) - (4) with zero-one costs (c_{ij}) is

$$L(K):\quad z(K) = \max \sum_{i=1}^{m} \max_{j=1,\,---,\,n} c_{ij}\, x_j$$
$$\sum_{j=1}^{n} w_j x_j \leq K$$
$$x \text{ binary}$$

Such problems are naturally associated with their set covering counterpart (15), for which (c_{ij}) is the incidence matrix, w the coefficient vector of the objective function.

Fisher [4] notes that if a set covering problem with incidence matrix (c_{ij}) has an optimal value s, the corresponding weighted plant location problem with weight limit s has an optimum z(s) = m. Furthermore, an optimal solution of the location problem L(s) is also optimal for the set covering problem.

The relationship between the two problems extends to their respective greedy heuristics. (The greedy heuristic for the set covering problem chooses the columns with the same criterion as the greedy location heuristic but stops only when all the rows of the problem are covered). In particular the columns of the greedy location heuristic solution coincide with the first columns of the set covering heuristic solution.

Chvatal [1] proves the following result for the greedy covering

Theorem (Chvatal [1]): For the set covering problem (15), with optimal value s, the value s_G of any greedy solution satisfies:

$$s_G \le s \sum_{j=1}^{d} \frac{1}{j}$$

where $d \le m$ is the maximum number of non-zero entries of any incidence column. The bound can be relaxed to

$$s_G \le s\,(1 + \ell n\, m)$$

The following property shows that, in the weighted case, the relaxation of Chvatal's bound can be inferred from the bound obtained for greedy location.

(Note that in the simple plant location problem with 0-1 costs:

$$z_R = 0$$

$$z_D \geq z \geq 0$$

therefore (7) implies the weaker bound (17) used below)

Property 1: The bound on the weighted plant location heuristic

(17) $$\frac{z_D - z_G}{z_D - z_R} \leq \prod_{h=1}^{H} \left(1 - \frac{w_h}{K}\right)$$

implies the weak bound on the set covering heuristic

(18) $$s_G \geq s\ (1 + \ell n\ m)$$

Proof: for the set covering problem defined in (15), consider the weighted plant location problem whose weight limit is s.

Apply the greedy location heuristic: z_G rows are covered, and the greedy covering heuristic would proceed to cover the remaining $z - z_G = m - z_G$ rows, which constitute a second covering problem, with optimum s_R and greedy value s_{RG}.

(18) holds naturally when m = 1, since the heuristic solution is always optimal.

Suppose it holds for $\ell = 1, ---, m - 1$, therefore for $\ell = m - z_G = z - z_G$

$$s_{RG} \leq s_R \left[1 + \ell n\ (z - z_G)\right]$$

obviously

$$s_R \leq s$$

and, by definition of the covering heuristic

$$s_G = \sum_{h=1}^{H} w_h + s_{RG}$$

therefore

$$s_G \leq \sum_{h=1}^{H} w_h + s \left[1 + \ell n\ (z - z_G)\right]$$

and, by (17), with K = s:

$$s_G \le \sum_{h=1}^{H} w_h + s\left[1 + \ell n\ z + \sum_{h=1}^{H} \ell n\left(1 - \frac{w_h}{s}\right)\right]$$

$$s_G \le s\ (1 + \ell n\, z) + \sum_{h=1}^{H} \left[w_h + s\,\ell n\left(1 - \frac{w_h}{s}\right)\right]$$

but each of the terms of the summation is non-positive, therefore

$$s_G \le s\ (1 + \ell n\, z)$$

This Property is a generalization of [4, Theorem 1] for the weighted case.

The following section introduces a new bound for the simple plant location greedy heuristic, which can be shown to be "asymptotically" equivalent to the bound proved by Cornuejols, Fisher and Nemhauser [2].

4. A new bound for the greedy location heuristic

Theorem 2: Consider the simple K-plant location problem (1) - (5) with zero-one costs c_{ij}. Let d be the maximum number of non-zero entries in a column of C, and k be the smallest number such that:

$$\frac{1}{d} + \frac{1}{d-1} + \text{---} + \frac{1}{k+1} \leq 1$$

then the value z_G of the greedy solution is related to the optimal value z of the problem by

$$\frac{z - z_G}{z} \leq \frac{k}{d}\left(\frac{1}{d} + \text{---} + \frac{1}{k+1}\right)$$

Proof: First we show that the bound is valid. When the greedy heuristic is applied, the values of the multipliers ρ_ℓ, ℓ = 1, ---, K, are non-increasing. Clearly $\rho_1 = d$. Let $h = \rho_K$. Define s_i to be the number of iterations during which the multipliers are equal to i, for i = d, ---, h. (Note that s_i may take the value 0.)

(19) $\quad s_d + s_{d-1} + \text{---} + s_h = K$

(20) $\quad z_G = ds_d + (d-1)s_{d-1} + \text{---} + hs_h.$

Note also that the following inequalities must hold.

(21.d) $\quad z \leq Kd$

(21.d-1) $\quad z \leq ds_d + K(d-1)$

(21.i) $\quad z \leq ds_d + (d-1)s_{d-1} + \text{---} + (i+1)s_{i-1} + Ki$

(21.h) $\quad z \leq ds_d + (d-1)s_{d-1} + \text{---} + (h+1)s_{h-1} + Kh$

The reason for which (21.i) must hold for $i = d, d-1, --- h$, is that an optimal solution can cover at most Ki of the rows left uncovered after $s_d + s_{d-1} + --- + s_{i-1}$ iterations of the greedy heuristic and at best it will cover also the rows covered by that partial greedy solution.

Divide through by z. Set $x_i = \frac{s_i}{z}$ for $i = d, d-1, ---, h$ and use equation (19) to get rid of $\frac{K}{z}$ whenever it appears in the constraint set (21). Then (19) - (21) yields

(22) $$\frac{z_G}{z} = dx_d + (d-1)x_{d-1} + --- + hx_h$$

(23.d) $$dx_d + dx_{d-1} + --- + dx_h \geq 1$$

(23.(d-1)) $$[d + (d-1)]x_d + (d-1)x_{d-1} + --- + (d-1)x_h \geq 1$$

(23.h) $$[d+h]x_d + [(d-1) + h]x_{d-1} + --- + hx_h \geq 1$$

(24) $$x_d, x_{d-1}, ---, x_h \geq 0$$

The minimum of (22) subject to (23) - (24) is bounded below by any feasible solution of the dual problem.

Max $$y_d + y_{d-1} + --- + y_h$$

(25.d) $$dy_d + [d + (d-1)]y_{d-1} + --- + (d+h)y_h \leq d$$

$$dy_d + (d-1)y_{d-1} + --- + [(d-1) + h] \leq d-1$$

(25.h) $$dy_d + (d-1)y_{d-1} + --- + hy_h \leq h$$

$$y_d, y_{d-1}, ---, y_h \geq 0$$

When $h \geq k$ (as defined in the statement of the theorem) it is easy to check that all the above constraints are satisfied by the solution

$$y_{h-i} = \frac{h}{(h+i)(h+i+1)} \quad \text{for } i = 0, ---, d-h-1$$

and

$$y_d = \frac{h}{d}\left[1 - \left(\frac{1}{d} + \frac{1}{d-1} + --- + \frac{1}{h+1}\right)\right].$$

[In fact every constraint (25) can be satisfied as an equality and the corresponding linear system can be diagonalized by subtracting (25.i) from (25.(i-1)) i = h+1, ---, d. This yields the above solution.].

Its value $y_d + y_{d-1} + --- + y_h$ is a lower bound on (22).

$$\frac{z_G}{z} \geq y_d + y_{d-1} + --- + y_h = \frac{h}{d}\left[1 - \left(\frac{1}{d} + --- + \frac{1}{h+1}\right)\right] + \sum_{i=0}^{d-h-1} \frac{h}{(h+i)(h+i+1)}.$$

Simple calculations show that

$$\sum_{i=0}^{d-h-1} \frac{h}{(h+i)(h+i+1)} = 1 - \frac{h}{d}$$

Thus

$$\frac{z_G}{z} \geq 1 - \frac{h}{d}\left(\frac{1}{d} + \frac{1}{d-1} + --- + \frac{1}{h+1}\right) \tag{26}$$

When $h < k$ define ℓ to be the smallest integer such that $\frac{1}{h+1} + \frac{1}{h+2} + --- + \frac{1}{d-\ell-1} + \frac{1}{d-\ell} < 0$. Then the following solution satisfies all the constraints (25) as well as nonnegativity.

$$y_d = --- = y_{d-\ell-1} = 0$$

$$y_{d-\ell} = \frac{h}{d-\ell}\left(1 - \left(\frac{1}{d-\ell} + \frac{1}{d-\ell-1} + --- + \frac{1}{h+1}\right)\right)$$

$$y_{h-i} = \frac{h}{(h+i)(h+i+1)} \quad \text{for } i = 0, ---, d-\ell-h-1.$$

Its value $y_d + y_{d-1} + --- + y_h = 1 - \frac{h}{d-\ell}\left(\frac{1}{d-\ell} + \frac{1}{d-\ell-1} + --- + \frac{1}{h+1}\right)$ is a lower bound on $\frac{z_G}{z}$.

(27) $\frac{z_G}{z} \geq 1 - \frac{h}{d-\ell}\left(\frac{1}{d-\ell} + \frac{1}{d-\ell-1} + --- + \frac{1}{h+1}\right).$

The smallest* bound in the set (26) - (27) is obtained when h = k (a simple fact to check). Thus

$$\frac{z_G}{z} \geq 1 - \frac{k}{d}\left(\frac{1}{d} + \frac{1}{d+1} + --- + \frac{1}{k+1}\right)$$

where now the bound is valid for any problem instance (independently of h). Thus

$$\frac{z-z_G}{z} \leq \frac{k}{d}\left(\frac{1}{d} + \frac{1}{d+1} + --- + \frac{1}{k+1}\right).$$

There remains to prove that the bound is attained. For any positive integer d, consider the simple plant location problem (1) - (5)

$m = d \,.\, d!$

$K = d!$

$n = 2K$

(c_{ij}) as shown in Figure 1

The first K columns are orthogonal to each other and can be subdivided into d subsets of columns, namely

first subset of	$\frac{d!}{d}$	columns, each containing d ones
second subset of	$\frac{d!}{d-1}$	columns, each containing d-1 ones
ℓ^{th} subset of	$\frac{d!}{d-\ell+1}$	columns, each containing d-ℓ+1 ones
$(d-k+1)^{th}$ subset of	$d!\left(1 - \frac{1}{d} - \ldots - \frac{1}{k+1}\right)$	columns, each containing k ones

The next K columns are formed by d superposed "identity" blocks. (This cost matrix is very similar to ([5], Figure 1), except that the first K columns are sufficient to constitute a greedy solution, whereas

$K\left(\frac{1}{d} + \frac{1}{d-1} + --- + \frac{1}{2} + 1\right)$ are needed for the greedy covering.)

---

*Note from J.-M. Thizy: 'largest' in the original Working paper.

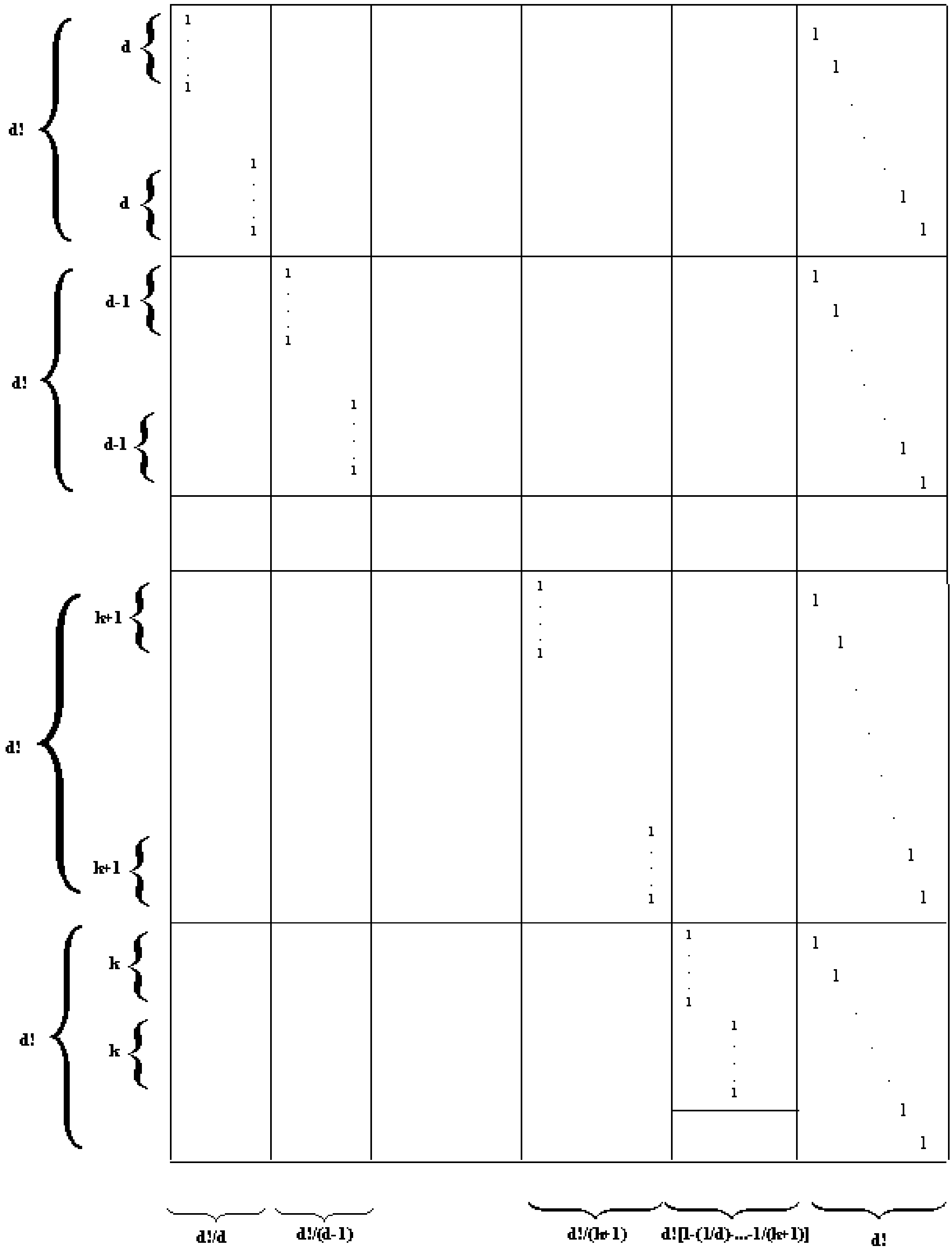

Figure 1.

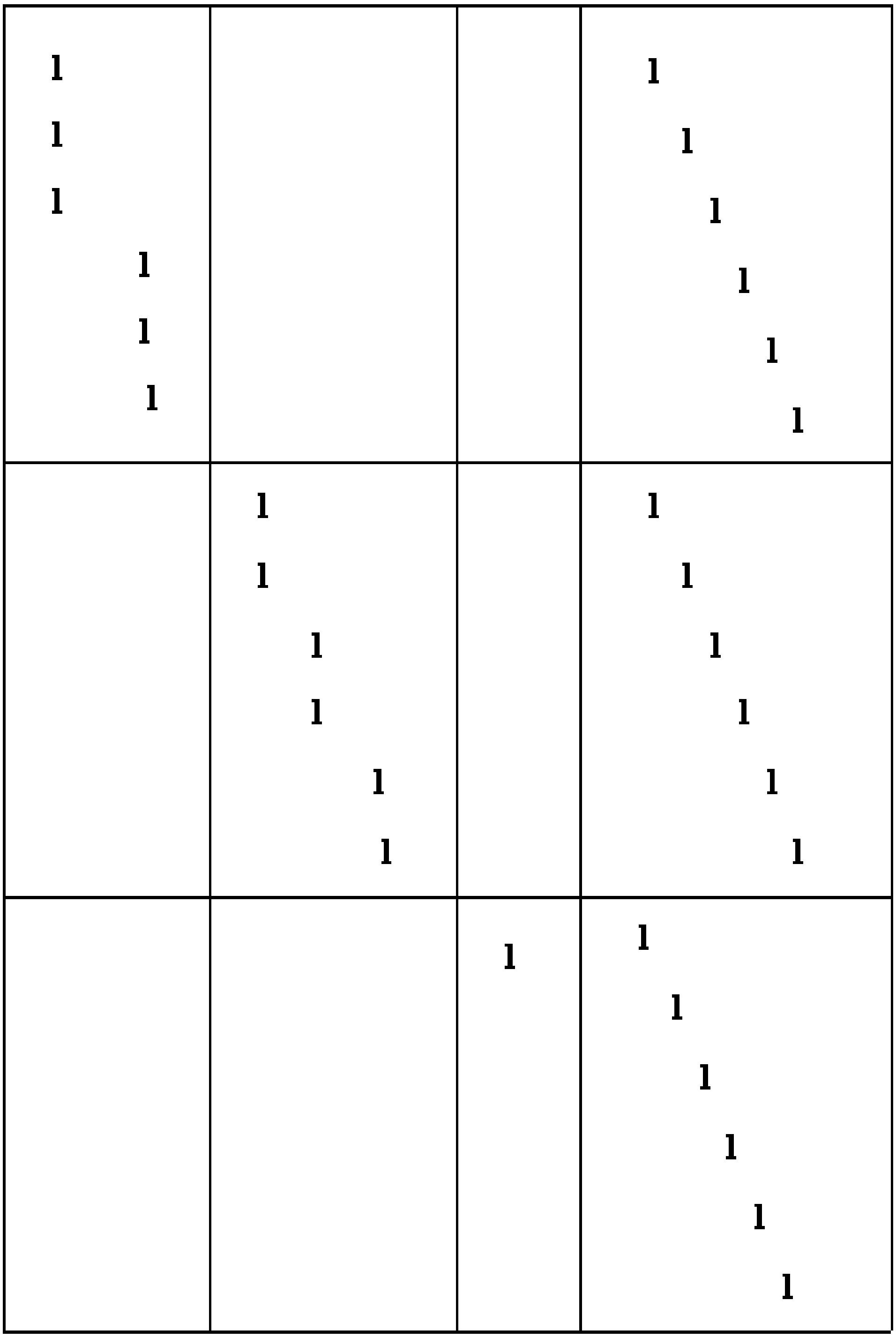

Figure 2. d = 3

The greedy solution consists of the first K columns; the following table tallies the number of row covered for each subset of columns chosen by the heuristic:

1^{st}	subset of $\frac{d!}{d}$	each containing d	ones: TOTAL d! rows
2^{nd}	subset of $\frac{d!}{d-1}$	each containing (d-1)	ones: TOTAL d! rows
ℓ^{th}	subset of $\frac{d!}{d-\ell+1}$	each containing $(d-\ell+1)$	ones: TOTAL d! rows
$(d-k+1)^{th}$	subset of $d!\left(1-\frac{1}{d}-\ldots-\frac{1}{k+1}\right)$	each containing k ones:	TOTAL $kd!\left(1-\frac{1}{d}-\ldots-\frac{1}{k+1}\right)$ rows

$$z_G = \text{TOTAL \# ROWS COVERED: } (d-k)d! + kd!\left(1-\frac{1}{d}-\ldots-\frac{1}{k+1}\right)$$

$$z_G = d!\left[d-k+k-k\left(\frac{1}{d}+---+\frac{1}{k+1}\right)\right]$$

$$\frac{z_G}{z} = \frac{d!}{d.d!}\left[d-k\left(\frac{1}{d}+---+\frac{1}{k+1}\right)\right] = 1-\frac{k}{d}\left(\frac{1}{d}+---+\frac{1}{k+1}\right)$$

$$\frac{z-z_G}{z} = 1-\frac{z_G}{z} = \frac{k}{d}\left(\frac{1}{d}+---+\frac{1}{k+1}\right)$$

The proof shows that if the greedy heuristic stops with $\rho_K = h < k$, then a tighter bound than that of Theorem 2 holds.

<u>Corollary 2:</u> Consider the problem (1) - (5) with zero-one costs c_{ij}.

$$\frac{z-z_G}{z} \le \frac{h}{d}\left(\frac{1}{d}+\frac{1}{d-1}+---+\frac{1}{h+1}\right)$$

where $h = \rho_K$ when the greedy heuristic terminates.

Using the canonical reduction of [3], Theorem 2 can be extended to a general problem (1) - (5) where the objective function can be any integral matrix C (since the greedy heuristic on the associated zero-one canonical problem chooses the same locations.)

<u>Corollary 3:</u> Consider the problem (1) - (5). Let $d = \max_{j=1,---,n} \sum_i c_{ij}$ and k the smallest integer such that:

$$\frac{1}{d-z_R} + \frac{1}{d-1-z_R} + --- + \frac{1}{k+1} \leq 1.$$

Then

$$\frac{z-z_G}{z-z_R} \leq \frac{k}{d-z_R}\left(\frac{1}{d-z_R} + \frac{1}{d-z_R-1} + --- + \frac{1}{k+1}\right).$$

<u>Property 2:</u> The bound on the simple plant location heuristic defined in Theorem 2:

$$(28)_d \qquad \frac{z-z_G}{z} \leq \frac{k}{d}\left(\frac{1}{d} + --- + \frac{1}{k+1}\right)$$

implies Chvatal's bound for the set covering heuristic

$$(29)_d \qquad s_G \leq s\left(\frac{1}{d} + \frac{1}{d-1} + --- + \frac{1}{2} + 1\right)$$

<u>Proof:</u> Consider the set covering problem (15) and the associated simple plant location problem (1) - (5) with K = s. Recall that d is the maximum number of ones contained in any column of (c_{ij}); in this proof, k is defined as the minimum integer such that:

$$\frac{1}{d} + \frac{1}{d-1} + --- + \frac{1}{k+1} \leq 1$$

The set covering greedy heuristic chooses the same first K columns as the location greedy heuristic.

a) If the K^{th} column covers d new rows, it yields an optimal covering solution, since any optimal covering has K = s columns, each column covering at most d rows.

Therefore $\quad s_G = s \leq s\left(\frac{1}{d} + \frac{1}{d-1} + --- + \frac{1}{2} + 1\right)$

b) If the first K columns of the greedy solution do not cover all the rows, covering the remaining rows constitutes a second (restricted) set covering problem, whose optimum s_R satisfies:

$$s_R \leq s$$

The sequel of the greedy heuristic ($(K+1)^{th}$, $(K+2)^{th}$, --- columns) coincides with the greedy heuristic of this restricted problem, whose value will be denoted s_{RG}, therefore

$$s_G = s + s_{RG} \tag{30}$$

We may also assume that Chvatal's bound $(29)_\delta$ holds for $\delta = 1, 2, ---, d-1$. (It does hold for $\delta = 1$, when every heuristic solution is optimal.)

c) If the $(K+1)^{th}$ column of the greedy covering covers h new rows, with $h < k$:

$$\frac{1}{d} + \frac{1}{d-1} + --- + \frac{1}{h+1} > 1 \tag{31}$$

since $h < d$, by the previous recursive assumption

$$s_{RG} \leq s_R \left(\frac{1}{h} + \frac{1}{h-1} + --- + 1\right)$$

$$s_{RG} \leq s \left(\frac{1}{h} + \frac{1}{h-1} + --- + 1\right) \tag{32}$$

(30) and (32) imply:

$$s_G \leq s + s \left(\frac{1}{h} + \frac{1}{h-1} + --- + 1\right)$$

and, by (31)

$$s_G \leq s \left(\frac{1}{d} + \frac{1}{d-1} + --- + \frac{1}{h+1}\right) + s \left(\frac{1}{h} + \frac{1}{h-1} + --- + 1\right)$$

which proves $(29)_d$.

d) Otherwise, if $k \leq h \leq d-1$

Let s_i be the number of columns chosen by the <u>restricted</u> greedy covering that cover i new rows (i = 1, 2, ---, h). (Note: s_h does not include any of the first K columns chosen before restricted problem)

Let $\qquad S_j = \sum_{i=1}^{j} s_i \quad j = 1, ---, h$

since $\quad h \leq d-1$, as for $(29)_j$:

$$(33)_j \qquad S_j \leq s\left(\frac{1}{j} + \frac{1}{j-1} + --- + 1\right) \quad \text{for } j = 1, ---, h$$

Show that:

$$(34)_\ell \qquad \sum_{j=1}^{l-1} S_j \leq s.\ell\left(\frac{1}{\ell} + \frac{1}{\ell-1} + --- + \frac{1}{3} + \frac{1}{2}\right) \qquad \ell = 2, ---, h + 1$$

This is true for $\ell = 2$:

$$S_1 \leq s \times 2 \times \frac{1}{2}$$

suppose it is true for i = 2,---, ℓ

then $$\sum_{j=1}^{l} S_j \leq S_\ell + s.\ell\left(\frac{1}{\ell} + \frac{1}{\ell-1} + --- + \frac{1}{3} + \frac{1}{2}\right)$$

and by $(33)_\ell$

$$\sum_{j=1}^{l} S_j \leq s\left(\frac{1}{\ell} + \frac{1}{\ell-1} + --- + 1\right) + s.\ell\left(\frac{1}{\ell} + \frac{1}{\ell-1} + --- + \frac{1}{3} + \frac{1}{2}\right)$$

$$\sum_{j=1}^{l} S_j \leq s + s(\ell+1)\left(\frac{1}{\ell} + \frac{1}{\ell-1} + --- + \frac{1}{3} + \frac{1}{2}\right)$$

$$\leq s(\ell+1)\left(\frac{1}{\ell+1} + \frac{1}{\ell} + --- + \frac{1}{3} + \frac{1}{2}\right)$$

which proves the property for $\ell+1$. Hence the property holds for $\ell = 2, ---,$ $h + 1$.

Also:

$$z - z_G = \sum_{i=1}^{h} i\, s_i \leq \frac{hz}{d}\left(\frac{1}{d} + \frac{1}{d-1} + \text{---} + \frac{1}{h+1}\right) \qquad \text{by} \quad (28)_d$$

but, since K = s: $z \leq s\, d$

therefore $\sum_{i=1}^{h} i\, s_i \leq s.h\left(\frac{1}{d} + \frac{1}{d-1} + \text{---} + \frac{1}{h+1}\right)$

Finally $s_G = s + s_{RG} = s + S_h$

and by simple algebraic manipulation $s_G = s + \dfrac{\sum_{i=1}^{h} is_i}{h} + \dfrac{\sum_{i=1}^{h-1} S_i}{h}$

using $(33)_h$ and $(34)_h$, we obtain

$$s_G \leq s + s\left(\frac{1}{d} + \frac{1}{d-1} + \text{---} + \frac{1}{h+1}\right) + s\left(\frac{1}{h} + \text{---} + \text{---} + \frac{1}{2}\right)$$

which in turn proves $(29)_d$.

REFERENCES

1. V. Chvatal, "A Greedy Heuristic for the Set-Covering Problem", Math. of Oper. Res. Vol. 4, No. 3, August 1979, pp. 233-235.

2. 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, Vol. 23, No. 8 (1977) 789-810.

3. G. Cornuejols, G.L. Nemhauser, and L.A. Wolsey, "A Canonical Representation of Simple Plant Location Problems and its Applications", SIAM Journal on Algebraic & Discrete Methods, 1, 261-272 (1980).

4. M.L. Fisher, "On the Equivalence of Greedy Covering and Greedy Location", The Wharton School, University of Pennsylvania Working Paper 79-11-01, October 1979.

5. D.S. Johnson, "Approximation Algorithms for Combinatorial Problems", J. Comput. System Sci. 9, 256-298 (1974).

6. L. Lovasz, "On the Ratio of Optimal Integral and Fractional Covers", Discrete Math. 13, 383-390 (1975).

7. L.A. Wolsey, "Maximizing Real-Value Submodular Functions: Primal and Dual Heuristics for Location Problems", Core Report, 1980.