AN EX-POST BOUND ON THE GREEDY HEURISTIC FOR THE UNCAPACITATED FACILITY LOCATION PROBLEM [*]

Jean-Michel Thizy[1]

Abstract. A bound for the greedy heuristic applied to the K-facility location problem can be calculated, using values gathered during the calculation of the heuristic. The bound strengthens a well-known bound for the heuristic. Computational experiments show that this bound can be beneficial when the number of facilities is small or close to the total number of potential sites. In addition, it is consistent with previous results about the influence of the data characteristics upon the optimal value.

1. Introduction

The uncapacitated facility location problem, also known as the simple plant location problem, is a centerpiece of logistic planning as well as combinatorial optimization:

"The simple plant location problem is one of the simplest mixed integer problems which exhibits all the typical combinatorial difficulties of mixed (0,1) programming and at the same time has a structure that invites the application of various specialized techniques" [15].

[9] offers a comprehensive study of its properties and applications. It is an
NP-hard problem for which many approximation algorithms have been designed.
This article focuses on a very well-known variety of the problem, the K-facility
location problem [4], with alternative names such as *p-plant location problem* or
p-facility location problem. The best-known approximation algorithm for this prob-
lem, the Greedy Heuristic [19] was proved never to deviate from the optimal solu-
tion by more than $\frac{1}{e}$ (in relative error), where e is the base of natural logarithms [4].
A new bound was proposed in [6, 23] for the K-facility location problem with 0-1
revenues, no fixed cost and extended by the author in an oral presentation [7]. The
derivation of the new bound is given here, along with an extension to the general
formulation and comparison with the previous bound; a computational assessment
of its performance shows that the bound adapts to the particularities of the data
of each problem instance. Thus, it presents an advantage over the classical bound
of $\frac{1}{e}$, especially when the number of facilities is either small or close to the total
number of potential sites. In addition, it is consistent with results about the in-
fluence of the data characteristics upon the gap between the optimal value of the
problem and that of its linear relaxation [1, 22].

The next section contains a formulation of the problem. Section 3 describes the
Greedy Heuristic. Section 4 analyzes a Lagrangian relaxation of the problem. Sec-
tion 5 offers a new bound on the performance of the Greedy Heuristic. Section 6
provides a computational assessment of this bound and Section 7 presents lines of
extension.

2. PROBLEM FORMULATION

The K-facility location problem P consists of selecting at most K facilities to
keep open among n potential sites in order to minimize the revenue of serving m
demands. The problem can be formulated as:

$$P: \quad \max \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} - \sum_{j=1}^{n} f_j y_j \tag{1}$$

$$\text{subject to:} \quad \sum_{j=1}^{n} x_{ij} = 1 \quad \text{for} \quad i = 1, ..., m \tag{2}$$

$$0 \le x_{ij} \le y_j \le 1 \quad \text{for} \quad i = 1, ..., m \quad \text{and} \quad j = 1, ..., n \tag{3}$$

$$\sum_{j=1}^{n} y_j \le K \tag{4}$$

$$y_j \quad \text{integer} \quad \text{for} \quad j = 1, ..., n \tag{5}$$

where:

y_j is 1 if a facility is opened at Location j and 0 otherwise;
x_{ij} is 1 if Location j satisfies Demand i and 0 otherwise;

c_{ij} is the revenue of satisfying Demand i from Location j;

f_j is the fixed cost of a facility at Location j (revenues or costs can be negative).

3. The greedy heuristic [19]

The greedy heuristic yields a feasible solution of the problem by choosing facilities one at a time; at each of its (K) iterations, the chosen facility is one that produces the largest improvement in the objective value. To formalize this method, the following algorithm is essentially reproduced from [4]; $H = k$ counts the number of iterations: the facilities chosen are recorded in J^*, $u_i^k = \max_{j \in J^*} c_{ij}$ denotes the opportunity revenues facing the decision makers, *i.e.*, the highest distribution revenues that can be generated from the incumbent set of facilities already selected. For each candidate location j, $\rho_j(u)$ aggregates over all demands the incremental revenue (above the opportunity revenues) that the facility would bring by operating there, less f_j plus $F = \max_{j=1,\ldots,n} f_j$. The Greedy Heuristic can be defined by the algorithm:

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

Step 2

    a. Let $\rho_j(u^k) = F - f_j + \sum_{i=1}^m \max\left(0, c_{ij} - u_i^k\right), j \notin J^*$.

    b. Find $j_k \notin J^*$ such that $\rho_k = \rho_{j_k}(u^k) = \max_{j \notin J^*} \rho_j(u^k)$.

    c. If $\rho_k < F$ and $|J^*| \geq 1$, set $\rho_k = F$ and go to step 4; otherwise set $J^* = J^* \bigcup \{j_k\}$.

    d. If $|J^*| = K$, go to step 4, else go to step 3.

Step 3   Set $k = k + 1$. For $i = 1, \ldots, m$, set $u_i^k = \max_{j \in J^*} c_{ij} = u_i^{k-1} + \max(0, c_{ij_{k-1}} - u_i^{k-1})$. Go to step 2.

Step 4   Stop; the greedy solution is given by $y_j = 1, j \in J^*$ and $y_j = 0$ otherwise, and its objective value is $z_G = \sum_{i=1}^m u_i^1 + \sum_{h=1}^k \rho_h - kF$.

4. A Lagrangian relaxation of the problem

It is well-known that the optimum value of the linear relaxation (1)–(4) $z = \min_u z(u)$, where $u = (u_1, \ldots, u_m)$ are multipliers for Constraints (2), and

$P(u) : z(u) = \max_{x,y} z(x, y, u)$, with

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

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

subject to (3)–(4).

$P(u)$ is a continuous 0–1 knapsack problem, for which an analytic upper bound is available. Define a permutation $\{\Im(u,j)\}_{j=1,...,n}$ that orders the values $\rho_j(u)$ nonincreasingly: $\rho_{\Im(u,k)} \geq \rho_{\Im(u,h)}$ if $k < h$; then for any vector u of multipliers:

$$z \leq z(u) \leq \sum_{i=1}^{m} u_i + \sum_{h=1}^{r(u)-1} \rho_{\Im(u,h)} + \rho_{r(u)} \left[K + 1 - r(u)\right] - KF \qquad (6)$$

where $r(u)$ satisfies: $r(u) - 1 \leq K \leq r(u)$, *i.e.* the optimum z can be bounded and one value of u will be shown to be yielded by the Greedy algorithm. The proof is incremental: first, Lemma 1 considers the important case where Constraint (4) must hold as an equality. Note then that step 2c of the Greedy algorithm can be simplified as:

c. Set $J^* = J^* \bigcup \{j_k\}$.

The original specification will be reinstated in Proposition 1.

5. WORST-CASE ANALYSIS OF A GREEDY ALGORITHM

Lemma 1. The value z_G of a greedy solution is related to the optimum z of the K-facility location problem relaxation (1)–(4) for which Constraint (4) must hold as an equality, $0 = \min_{j=1,...,n} c_{ij}$ for $i = 1,...,m$ and $f_j = 0$ for $j = 1,...,n$ by:

$$\frac{z - z_G}{z} \leq \frac{\rho_{j_K}}{\rho_{j_{\underline{\kappa}}}} \left(1 - \frac{\rho_{j_{\underline{\kappa}+1}}}{\rho_{j_{\underline{\kappa}}}} + ... + 1 - \frac{\rho_{j_K}}{\rho_{j_{K-1}}}\right) \qquad (7)$$

where $\underline{\kappa}$ is the smallest index such that $1 - \frac{\rho_{j_{\underline{\kappa}+1}}}{\rho_{j_{\underline{\kappa}}}} + 1 - \frac{\rho_{j_{\underline{\kappa}+2}}}{\rho_{j_{\underline{\kappa}+1}}} + ... + 1 - \frac{\rho_{j_K}}{\rho_{j_{K-1}}} \leq 1$.

Proof. Let s_i be the number of iterations during which the increment is equal to ρ_i, for $i = 1,...,s_1, 1 + s_1,...,s_1 + s_2, ..., s_1 + s_2 + ... + s_h$ (assuming without loss of generality that $h \geq 4$). Note that s_i may be 0. Define $\varrho_i = \rho_{s_i}$ and η by $\varrho_\eta = \rho_{j_{\underline{\kappa}}}$.

$$s_1 + s_2 + ... + s_h = K \qquad (8)$$

$$z_G = \varrho_1 s_1 + \varrho_2 s_2 + ... + \varrho_h s_h \qquad (9)$$

$$z \leq K\varrho_1 \qquad (10.1)$$

$$z \leq \varrho_1 s_1 + K\varrho_2 \qquad (10.2)$$

$$...$$

$$z \leq \varrho_1 s_1 + \varrho_2 s_2 + ... + \varrho_{i-1} s_{i-1} + K\varrho_i \qquad (10.i)$$

$$...$$

$$z \leq \varrho_1 s_1 + \varrho_2 s_2 + ... + \varrho_{h-1} s_{h-1} + K\varrho_h. \qquad (10.h)$$

Divide through by z. Set $x_i = \frac{s_i}{z}$ for $i = 1, 2, ..., h$ and use Equation (8) to eliminate $\frac{K}{z}$ whenever it appears in the constraint set (10). Then (8)–(10) yields

$$\frac{z_G}{z} = \varrho_1 x_1 + \varrho_2 x_2 + ... + \varrho_h x_h \tag{11}$$

$$\varrho_1 x_1 + \varrho_1 x_2 + ... + \varrho_1 x_h \geq 1 \tag{12.1}$$

$$(\varrho_1 + \varrho_2) x_1 + \varrho_2 x_2 + ... + \varrho_2 x_h \geq 1 \tag{12.2}$$

$$...$$

$$(\varrho_1 + \varrho_i) x_1 + (\varrho_2 + \varrho_i) x_2 + ... + (\varrho_{i-1} + \varrho_i) x_{i-1} + \varrho_i x_i \geq 1 \tag{12.i}$$

$$...$$

$$(\varrho_1 + \varrho_h) x_1 + (\varrho_2 + \varrho_h) x_2 + ... + (\varrho_{h-1} + \varrho_h) x_{h-1} + \varrho_h x_h \geq 1 \tag{12.h}$$

$$x_1, x_2, ..., x_h \geq 0. \tag{13}$$

A feasible solution of the system (12)–(13) is:

$$x_i = 0 \quad \text{for} \quad i = 1, ..., \eta - 1 \tag{14.1}$$

$$x_i = \frac{1}{\varrho_\eta} \left(1 - \frac{\varrho_{i+1}}{\varrho_i} \right) \quad \text{for} \quad i = \eta, ..., h - 1 \tag{14.2}$$

and

$$x_h = \frac{1}{\varrho_\eta} \left[1 - \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \right]. \tag{14.3}$$

To prove that it is optimal for the minimizing objective function (11), a feasible solution of its dual problem below is calculated afterward.

Max $y_1 + y_2 + ... + y_h$

s.t. $\varrho_1 y_1 + (\varrho_1 + \varrho_2) y_2 + ... + (\varrho_1 + \varrho_h) y_h \leq \varrho_1$

$\varrho_1 y_1 + \varrho_2 y_2 + (\varrho_2 + \varrho_3) y_3 + ... + (\varrho_2 + \varrho_h) y_h \leq \varrho_2$

$...$

$\varrho_1 y_1 + ... + \varrho_i y_i + (\varrho_i + \varrho_{i+1}) y_{i+1} + ... + (\varrho_i + \varrho_h) y_h \leq \varrho_i$

$....$

$\varrho_1 y_1 + \varrho_2 y_2 + ... + \varrho_h y_h \leq \varrho_h$

$y_1, y_2, ..., y_h \geq 0.$

A dual feasible solution that satisfies complementary slackness with the primal one, is:

$$y_i = 0 \quad \text{for} \quad i = 1, ..., \eta - 1, \tag{15.1}$$

$$y_i = \varrho_h \left(\frac{1}{\varrho_i} - \frac{1}{\varrho_{i-1}} \right) \quad \text{for} \quad i = \eta + 1, ..., h \tag{15.2}$$

and

$$y_\eta = \frac{\varrho_h}{\varrho_\eta} \left[1 - \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \right]. \tag{15.3}$$

This solution yields an objective value: $\zeta = 1 - \frac{\varrho_h}{\varrho_\eta} \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right)$ (see Appendix 1 for detailed calculations). The original notation ρ_j provides Bound (7). □

To fully define Bound (7) in the trivial case where $K = 1$, set $\rho_{j_o} = 1$ or any arbitrary non-zero value. Also, adapting the bound from the proof to the simpler expression (7) supposes that a ratio of two values $\rho_i = 0$ and $\rho_{i+1} = 0$ is 1. Define $z_R = \sum_{i=1}^m u_i^1 - K \max_{j=1,...,n} f_j$; this value represents a lower estimate of the worst possible choice of facilities, *i.e.* a selection with not even a simple calculation. It serves as a benchmark for the performance of the greedy heuristic which performs myopic calculations.

Proposition 1. The value z_G of a solution given by the Greedy Heuristic terminating in H iterations is related to the optimum z of the K-facility location problem relaxation (1)–(4) [for which Constraint (4) may be required to hold as an equality] by:

$$\frac{z - z_G}{z - z_R} \le \frac{\rho_{j_H}}{\rho_{j_\hbar}} \left(1 - \frac{\rho_{j_\hbar+1}}{\rho_{j_\hbar}} + ... + 1 - \frac{\rho_{j_H}}{\rho_{j_{H-1}}} \right) \tag{16}$$

where $\hbar$ is the smallest index such that $1 - \frac{\rho_{j_\hbar+1}}{\rho_{j_\hbar}} + 1 - \frac{\rho_{j_\hbar+2}}{\rho_{j_\hbar+1}} + ... + 1 - \frac{\rho_{j_H}}{\rho_{j_{H-1}}} \le 1$ and problem instances exist where (16) holds as an equality.

Proof.
1. First, consider the case of a problem P where Constraint (4) must hold as an equality and $f_j = 0$ for $j = 1, ..., n$. If $0 \ne \min_{j=1,...,n} c_{ij}$ for some demand $i \in \{1; ...; m\}$, Bound (7) holds for a problem P' derived from P *via* the modified revenues: $c'_{ij} = c_{ij} - u_i^1$ for $i = 1, ..., m$ and $j = 1, ..., n$. Note that $z' = z - z_R$ and $z'_G = z_G - z_R$, allowing us to assert that Problem P satisfies Proposition 1.
2. Next, consider a problem P where Constraint (4) must hold as an equality, $f_j = f_1$ for $j = 2, ..., n$ and compare this problem with a problem P' satisfying Case 1 for which $c'_{ij} = c_{ij}$ for $i = 1, ..., m$ and $f'_j = 0$ for $j = 1, ..., n$. Note that $z'_R = z_R + K f_1, z' = z + K f_1$ and $z'_G = z_G + K f_1$, allowing us to assert that Problem P satisfies Proposition 1.

3. Consider now a problem P where Constraint (4) must hold as an equality and the data have arbitrary values, in particular not satisfying the two cases analyzed before. Select an index k such that $f_k < \max_{j=1,...,n} f_j$. Compare Problem P with a problem P' for which $c'_{ij} = c_{ij}$ for $i = 1,...,m$ and $j = 1,...,n, f'_j = f_j$ for $j \neq k$ and $f'_k = \max_{j=1,...,n} f_j$. Include also a new demand $m + 1$ such that $c'_{m+1,j} = 0$ for $j \neq k$ and $c'_{m+1,k} = -f_k + \max_{j=1,...,n} f_j$. To any given optimal value $(x_{ij}, y_j)_{i=1,...,m;j=1,...,n}$ of Problem P, one can associate a similar optimal value $(x'_{ij} = x_{ij}, y'_j = y_j)_{i=1,...,m;j=1,...,n}$ and an assignment for Demand $m + 1$ satisfying Constraint (2) via $x'_{m+1,k} = y'_k$ and $x'_{m+1,j} \leq y_j$ for all other indices $j \neq k$. Note that $z'_R = z_R, z = z'$ and $z'_G = z_G$ so that Problem P satisfies Proposition 1 if Problem P' does. Now apply the procedure just described to Problem P' at most $n - 2$ times, until one obtains a problem $P^{(n-1)}$ such that $f_j^{(n-1)} = f_1^{(n-1)}$ for $j = 2,..., n$, which is Case 2, shown to satisfy Proposition 1.

4. For the original Constraint (4), let $p = \max\{1; \text{card}(\{j \in \{1,...,n\} \mid f_j < 0\})\}$. If $p \geq K$, then exactly K facilities are chosen both heuristically and optimally and therefore Constraint (4) holds exactly, which is Case 3. Else, associate the following problem P' satisfying Constraints (1)–(3) & Constraint (4) as an equality, with m demands and $n + K - p$ facilities: $c'_{ij} = c_{ij}$ for $i = 1,...,m, f'_j = f_j$ for $j = 1,...,n, c'_{ij} = \min_{j=1,...,n} c_{ij}$ for $i = 1,...,m$ and $f'_j = 0$ for $j = n+1, n+2, ..., n+K-p$. First note that $z'_R = z_R$. If the greedy heuristic for Problem P yields $\sum_{j=1}^{n} y_j = k < K$, then the greedy heuristic for Problem P' selects the additional facilities $n+1$, $n+2$, ..., $n+K-k$ and $z'_G = z_G$. The value z of an optimum solution of P is also an optimum for P', since no assignment of value for $y_j, j = n+1, n+2, ..., n+K-k$ can improve upon it. Therefore any such solution with $\sum_{j=1}^{n} y_j = k < K$ can be completed as an optimum solution of P' by setting $y'_j = y_j$ and $x'_{ij} = x_{ij}$ for $i = 1,...,m, j = 1,...,n$ and selecting values of y_j for $j = n+1, n+2, ..., n+K-k$ such that Constraint (4) is satisfied as an equality (with $x'_{ij} = 0$ for $i = 1,...,m$). Instances where (16) holds as an equality are given in Corollary 2.

□

When the Greedy Heuristic stops with $\rho_H = 0$, even when $\sum_{j=1}^{n} y_j = k < K$, then the value of the variables defined by the Greedy Heuristic as well as dual multipliers $u, \rho(u)$ and $-F$ for Constraint (4) defines a complementary pair: the greedy solution is optimal and Proposition 1 is naturally satisfied. Not only does Proposition 1 extend Lemma 1, it actually subsumes it since, under the premises of the latter, $z_R = 0$. Proposition 1 enables us to compare Bound (16) with the classical result about the K-facility location problem and more generally to link the result with the abundant literature about the uncapacitated facility location problem (1)–(5) alluded to in the introduction, as a special case of the K-facility location problem, when $K = n$.

Next, two relaxations of Proposition 1 are presented as corollaries. Each entails fewer parameters than Proposition 1 and instances exist where for any value of the parameter, their respective bound is reached exactly. In both cases, the parameters can be found before applying the heuristic, so they can be called a-priori bound in contrast with Bound (16). Each can be calculated easily. Next, Proposition 2 will help us establish Corollary 1.

Proposition 2. For any integer values of K and κ such that: $K \geq \kappa > 0$:

$$\left(\frac{K-1}{K}\right)^K \geq \left(\frac{K-\kappa}{K+1-\kappa}\right)^{K-\kappa+1}$$

$$= \max_{\rho_{j_\kappa} \geq \rho_{j_\kappa+1} \geq \dots \geq \rho_{j_K}} \frac{\rho_{j_K}}{\rho_{j_\kappa}} \left(1 - \frac{\rho_{j_\kappa+1}}{\rho_{j_\kappa}} + \dots + 1 - \frac{\rho_{j_K}}{\rho_{j_{K-1}}}\right)$$

Proof. The first inequality comes from the monotonicity of the sequence $\left(\frac{K-1}{K}\right)^K$, $K = 1, 2, \dots$ (see [4], p.797). The equality is proved in Appendix 2. $\square$

Proposition 2 offers an envelope for Bound (16), *i.e.* an overarching bound that depends on the parameter K only. This simplified bound is precisely the one proposed by [4] which can now be presented as a consequence of Proposition 1, as formalized in the following:

Corollary 1 [4]. The value z_G of a greedy solution is related to the optimum z of the K-facility location problem relaxation (1)–(4) by:

$$\frac{z - z_G}{z - z_R} \leq \left(\frac{K-1}{K}\right)^K \left(< \frac{1}{e}\right) \tag{17}$$

and, for any value of K, there exist data sets for which this bound is reached exactly.

The reader is directed to the original proof, corrected in [5] or to its variant [3]. However, the power of Proposition 1 is to yield other important consequences, such as Corollary 2, unpublished and therefore detailed next, with special attention given to the restrictive setting in which it also holds directly as a consequence of Lemma 1.

Corollary 2 [6]. Consider the K-facility location problem relaxation (1)–(4) [for which Constraint (4) may hold as an equality] with zero-one revenues c_{ij} and zero fixed costs f_j. Let d be the maximum number of non-zero entries in a column of $C = (c_{ij})_{i=1,\dots,m,\ j=1,\dots,n}$ and δ be the smallest number such that:

$$\frac{1}{\delta+1} + \dots + \frac{1}{d-1} + \frac{1}{d} \leq 1$$

then the value z_G of the greedy solution is related to the value z of the relaxation (1)–(4) by:

$$\frac{z - z_G}{z - z_R} \leq \frac{\delta}{d} \left(\frac{1}{\delta + 1} + ... + \frac{1}{d} \right) \qquad (18)$$

and, for any value of d, there exist data sets for which this bound is reached exactly when K facilities have been selected greedily.

Proof. In Lemma 1, the (possibly vanishing) variables s and therefore x were defined as functions of (pre-defined) ρ_j, for instance $\rho_j = 1, 2, ...d$. For any index j such that $\rho_j = k$, the algebraic simplification $1 - \frac{\rho_j}{\rho_{j+1}} = \frac{1}{k}$ shows that Bound (18) derives from Bound (7), as detailed formally in Appendix 3, which pinpoints the role of H.

A case of tight bound, provided in [6], is excerpted below, with zero fixed costs: K facilities are selected greedily. For any positive integer d, consider the K-facility location problem relaxation (1)–(4):

$m = d. \, d!$
$K = d!$
$n = 2K$
(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

...

pth subset of $\frac{d!}{d-p+1}$ each containing $(d - p + 1)$ ones

...

$(d - \delta + 1)$th subset of $d! \left(1 - \frac{1}{d} - ... - \frac{1}{\delta+1} \right)$ columns, each containing δ ones.

The next K columns are formed by d superposed "identity" blocks. (This revenue matrix is very similar to Figure 1 of [16], except that the first K columns are sufficient to constitute a greedy solution, whereas $K(\frac{1}{d} + \frac{1}{d-1} + \text{- - -} + \frac{1}{2} + 1)$ are needed for the greedy covering); Figure 2 provides an instance of the general case of Figure 1 for $d = 3$. It gives us an intuition about a regular pattern of row covering made up by the last 6 columns that offer an optimal solution, as opposed to a greedy choice of the first six columns (perturbation of real-value coefficients would provide tie-breaking).

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

first subset of $\frac{d!}{d}$ columns, each containing d ones: TOTAL d! rows

second subset of $\frac{d!}{d-1}$ columns, each containing $d - 1$ ones: TOTAL d! rows

...

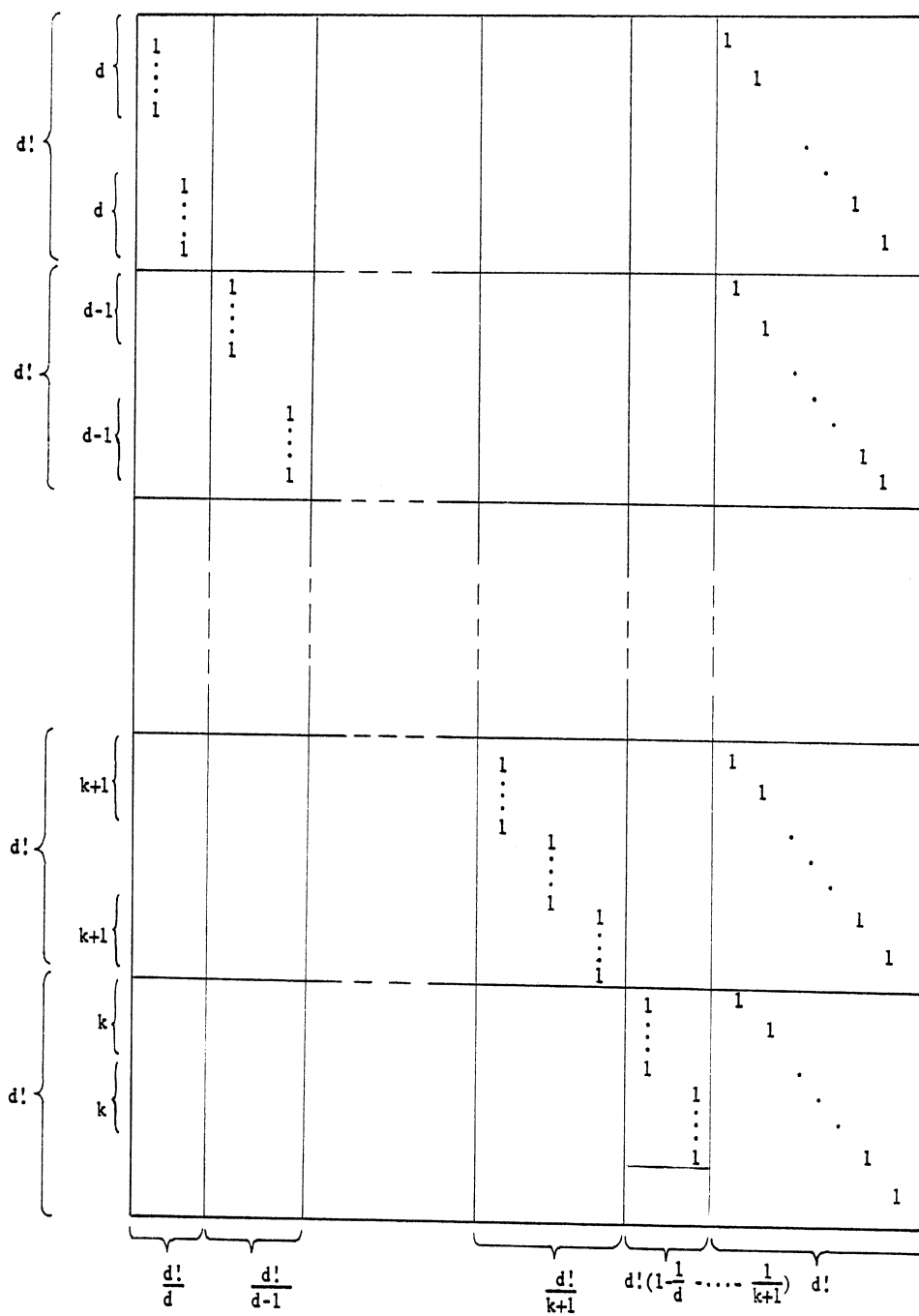pth subset of $\frac{d!}{d-p+1}$ each containing $(d - p + 1)$ ones: TOTAL d! rows

...

J.-M. THIZY

FIGURE 1

$(d - \delta + 1)$th subset of $d! \left(1 - \frac{1}{d} - ... - \frac{1}{\delta + 1}\right)$ columns, each containing δ ones:

$$\text{TOTAL} \quad \delta d! \left(1 - \frac{1}{d} - ... - \frac{1}{\delta + 1}\right) \text{rows}$$

$$z_G = \text{TOTAL NUMBER OF ROWS COVERED: } (d - \delta)d!$$

$$+ \delta d! \left(1 - \frac{1}{d} - ... - \frac{1}{\delta + 1}\right)$$

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

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

$$= 1 - \frac{\delta}{d} \left(\frac{1}{d} + ... + \frac{1}{\delta + 1}\right)$$

$$\frac{z - z_G}{z} = 1 - \frac{z_G}{z} = \frac{\delta}{d} \left(\frac{1}{d} + ... + \frac{1}{\delta + 1}\right). \qquad \qquad \square$$

Whereas Proposition 1 provides a bound tighter that those of Corollaries 1 and 2 (albeit one needing more information), neither bound introduced in Corollary 1 or 2 subsumes the other. This can be easily shown by considering Figure 2 where Bound (17) is 0.335 and Bound (18) is 0.278 and the example denoted C[22] of ([4], Figure 2), fully depicted by Figure 3, plus the specification $f = 0$, $K = 2$ and $d = 3$: Bound (17) is 0.25 and Bound (18) is 0.278.

6. COMPUTATIONAL VALIDATION OF THE NEW BOUND

As a reference, Bound (17) is calculated in Table 1 for various values of K (figures in the body of each subsequent table are percent values, except Tab. 13). Note the convergence toward $\frac{1}{e}$. Table 1 shows that Bound (17) is tighter for small values of K (of course, for $K = 1$, it is 0 and indeed the greedy heuristic yields an optimal solution). In this line, the subsequent analysis focuses on small values of K.

TABLE 1. Percent values of Bound (17).

K	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$(1\text{-}1/\text{K})^\text{K}$	25	30	32	33	33	34	34	35	35	35	35	35	35	36

Bound (17) is now compared numerically with Bound (16). The tests follow a standard suite using data sets which all comprise as many potential sites as demands. The first group, reported in Table 2, features major US cities from two sources: 30 cities in [10], 33 and 57 cities in [17] and also 100 points randomly selected in a rectangle in [18]. These problems, solved in [4, 8, 11, 21, 22], are deemed

J.-M. THIZY

1
1
1
            1
            1
            1

                                    1
                                        1
                                            1
                                                1
                                                    1
                                                        1

            1
            1
                1
                1
                    1
                    1

                                    1
                                        1
                                            1
                                                1
                                                    1
                                                        1

                        1

                                    1
                                        1
                                            1
                                                1
                                                    1
                                                        1

FIGURE 2. d = 3.

1	1	0	0
1	0	1	0
1	0	0	1
0	1	1	0
0	1	0	1
0	0	1	1

FIGURE 3. Example C^{22} of [4], Figure 2.

to be good representatives of K-facility location problems. Table 2 presents results consistent with Table 1. The tightening of Bound (16) is significant for small values of K and seems to be slightly enhanced by a larger problem size (which is summarized by n, the equal number of locations and demands). Deviations between actual optimal values and the greedy heuristic are much smaller than values for Bound (16) and largely uncorrelated with them; they are presented and discussed in Appendix 4. The general performance is confirmed by more extensive experimenting, using other randomly generated data tested in [1, 22] (the first reference contains original details about the random generation). Among these, the first set involves 4 series indexed by $p = 1$, 2, 3 and 4, each series p featuring 50 squares with a side of length 10^p, where each square contains 100 points connected by integer Euclidian distances. Table 3 is consistent with the experimental results of [1, 22] that noted the small influence of the length of the sides of the squares, *i.e.* the rounding of the data. Additional tests on this data set not reported here indicate that the range of bounds does not vary widely with the number of nodes, as will be shown on other data sets.

TABLE 2. Percent values of Bound (16) for standard data sets.

n \ K	2	3	4	5	6	7	8	9	10	11	12
30	16	25	20	21	28	28	28	28	29	30	30
33	20	22	25	28	28	31	32	32	32	32	32
57	11	23	24	24	31	31	31	31	31	31	31
100	14	18	22	22	22	26	32	32	33	32	33

A similar behavior can again be found in subsequent tables. The second set of data is based on averages of 50 random trees generated with $n = 50$, 100 or 150 nodes, using the method described in [12]. The distance between each pair of nodes is calculated as the length of the path joining them. Each row of Table 4 contains results for tree edge lengths equal to 1; for $K = 2$, the value of Bound (16) appears to improve with the number of nodes, but not to an extent that would permit to say that K scales proportionally with n. For higher values of K, no trend

TABLE 3. Percent values of Bound (16) for complete Euclidian graphs.

p \ K	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	11	17	24	29	29	28	28	29	30	30	31	31	32	32
2	11	17	25	30	28	28	28	29	29	30	31	32	32	32
3	11	17	25	30	29	28	28	29	30	31	31	32	33	33
4	11	17	25	30	28	28	28	29	30	31	31	32	33	33
5	11	17	25	30	28	28	28	29	30	31	31	32	33	33

is noticeable. It is worth noting that the experimental results of [1,22] found larger problems to display better linear relaxations. Our limited experiment reported in Table 14 does not show that the ratio $\frac{z-z_G}{z-z_R}$ decreases with the size of the graph.

TABLE 4. Percent values of Bound (16) for trees with unit revenues.

n \ K	2	3	4	5	6	7	8	9	10	11	12	13	14	15
50	10	18	25	27	28	29	30	30	30	30	31	30	30	30
100	8	19	25	27	28	29	30	31	31	31	32	32	32	32
150	8	19	25	27	28	30	30	31	31	32	32	32	32	33

Table 5 presents results for random nodes in the unit square, using the generation explained before. The length of an edge in the tree is the Euclidean distance between its endpoints rounded to an integer. A comparison between Tables 4 and 5 confirms that the value of Bound (16) does not appear to be influenced by the type of distance function, as found by the experimental results of [1,22].

TABLE 5. Percent values of Bound (16) for Euclidian trees.

n \ K	2	3	4	5	6	7	8	9	10	11	12	13	14	15
50	10	21	25	27	27	29	30	31	31	32	32	32	33	33
100	9	19	24	26	28	29	30	31	31	32	32	32	32	33
150	8	20	25	27	28	29	30	31	31	31	31	32	33	33

The third set of results relies on averages of 50 sequences of sparse graphs, starting from a random tree with 50 nodes, then adding at each q-th iteration 50 random edges to the graph of the previous iteration. Each row of the table corresponds to one value of q. For edge lengths equal to 1, Table 6 confirms remarks about earlier tables. For small values of K and low densities, Bound (16) increases with the density of the graph. This trend appears to be reversed as

the density grows, *i.e.* in lower rows of the table. However, this could also be a consequence of the fact that the distance between each pair of nodes decreases toward zero as the graph becomes fully dense.

TABLE 6. Percent values of Bound (16) for sparse graphs with unit revenues.

K / q	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	10	18	25	27	28	29	30	30	30	30	31	30	30	30
2	13	23	27	28	29	30	30	31	31	31	31	31	30	29
3	15	23	28	29	30	30	31	30	30	28	27	27	27	27
4	17	21	28	30	30	30	28	27	26	26	26	26	26	26
5	16	21	28	30	29	26	24	24	24	24	24	24	24	24
6	18	23	29	29	26	24	24	24	24	24	24	24	24	24
7	22	24	28	27	24	23	23	23	23	23	23	23	23	23
8	24	25	26	23	22	22	22	22	22	22	22	22	22	22
9	25	26	25	21	21	21	21	21	21	21	21	21	21	21
10	25	25	23	23	23	23	23	23	23	23	23	23	23	23
11	24	24	19	21	21	21	21	21	21	21	21	21	21	21
12	23	22	18	19	19	19	19	19	19	19	19	19	19	19
13	22	20	19	19	19	19	19	19	19	19	19	19	19	19
14	21	18	21	21	21	21	21	21	21	21	21	21	21	21
15	20	14	20	20	20	20	20	20	20	20	20	20	20	20
16	18	12	15	15	15	15	15	15	15	15	15	15	15	15
17	15	11	12	12	12	12	12	12	12	12	12	12	12	12

The remarks made about Tables 2–5 also apply to sparse Euclidian graphs. In Table 7, the nodes are in the unit square as in Table 5 and adjacent nodes are also paired by a Euclidian distance. The value of Bound (16) has decreased with respect to the trees of Table 6. On the other hand, even though the distances between nodes do not vanish as the graph density grows, Bound (16) does decrease, perhaps because the values of ρ_j decrease more regularly when the objective coefficients are less varied, as is the case of a dense graph endowed with the Euclidian property.

However, this last remark appears to be at odds with the results of Table 3, where revenue graininess did not affect the results. Therefore, a last set of data is generated with random integer revenues. For each value of $r = 1, 2$ and 3, a set S_r of 10 graphs with 100 nodes features revenues uniformly distributed in the range $[0, 10^r - 1]$. The results are in contrast with those of the preceding tables, as the values of Bound (16) are markedly below those of preceding tables, especially for small values of K. On the other hand, as shown in Table 3, Bound (16) shows

TABLE 7. Percent values of Bound (16) for sparse Euclidian graphs.

q \ K	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	10	21	25	27	27	29	30	31	31	32	32	32	33	33
2	11	22	27	28	30	30	31	31	32	32	33	33	33	33
3	12	23	27	28	30	31	31	32	32	32	33	33	33	33
4	13	22	27	28	29	30	31	31	32	32	32	33	33	33
5	13	22	27	28	29	30	31	31	32	32	32	32	32	32
6	13	21	27	28	29	30	31	31	32	32	32	32	33	33
7	14	21	27	28	29	30	31	31	31	32	32	32	32	32
8	13	21	26	28	29	30	30	31	31	31	32	32	32	32
9	13	20	27	28	29	29	30	30	31	31	32	32	32	32
10	13	19	26	28	29	30	30	31	31	31	31	32	32	32
11	13	19	27	29	29	30	30	30	31	31	31	31	32	32
12	13	20	26	28	29	29	29	30	30	30	31	31	31	31
13	13	20	26	29	29	29	29	30	30	30	31	31	31	31
14	13	19	26	28	29	29	29	30	30	31	31	31	31	31
15	13	18	26	29	29	29	29	30	30	30	31	31	31	31

TABLE 8. Percent values of Bound (16) for random revenues.

r \ K	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	22	25	25	28	29	30	31	30	30	30	31	29	29	19
2	21	25	26	28	29	30	30	31	31	32	32	32	32	33
3	21	25	26	28	29	30	30	31	31	32	32	32	32	33

little sensitivity to the range of the revenues c_{ij}, *i.e.* the effect of data rounding (which again evokes the experimental results of [1, 22]). Unreported results show, as Tables 4 and 5 did, that the values of Bound (16) do not vary significantly with the number of nodes.

Overall, the results show that the improvement brought by Bound (16) is particularly important for small values of K.

7. CONCLUSION AND FUTURE DIRECTIONS

In practice, Proposition 1 provides decision-makers with a better guarantee that the Greedy Heuristic solution is close to the optimum than was known before. It

also holds potential to improve our understanding of the relationship between performance guarantees for a more general class of heuristics.

– The relationship between (set) covering and (uncapacitated facility) location problems is well-known and its implication about the greedy heuristics for both classes of problems established in [14]. However, the latter treats a relaxation of Bound (17), which it proves to imply a relaxation of the bound for set-covering [2]. Following a special case [6] of 0–1 objective coefficients, [24] shows that Bound (18) implies the bound for set-covering [2]. Further work is likely to expand the class of covering and location problems that are related via Bound (16) and in particular, Bound (18). To this end, Formulation (8)–(10) can be expanded to one for submodular functions, much in the vein of Formulation (Q^I) for the submodular set covering problem [25] which yields a diversified bound for covering, reminiscent of Bound (16).

– The title of the article: ex-post bound and much of its focus highlight an algorithmic viewpoint, *i.e.* the stepwise construction of the Greedy Heuristic solution. A starting point for further research, consistent with [13, 20], is to view the locations chosen as a set, where the parameters ρ_j represent a simple data arrangement.

APPENDIX 1 - SOLUTION OF PROGRAM (12.1)–(12.h)–(13)

We show that the values of x defined in (14.1)–(14.3) yield a solution of Program (12.1)–(12.h)–(13).

For a row $i \leq \eta$:

$$\frac{\varrho_i}{\varrho_\eta} \left[1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} + 1 - \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 \right. \right.$$
$$\left. \left. - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \right] = \frac{\varrho_i}{\varrho_\eta} \geq 1.$$

Note that the inequality holds as an equality for $i = \eta$.

For a row $i > \eta$:

$$\frac{1}{\varrho_\eta} \left\{ \varrho_\eta \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} \right) + \varrho_{\eta+1} \left(1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} \right) + ... + \varrho_{i-1} \left(1 - \frac{\varrho_i}{\varrho_{i-1}} \right) \right.$$
$$+ \varrho_i \left[1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} + 1 - \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 \right. \right.$$
$$\left. \left. \left. - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \right] \right\} = \frac{1}{\varrho_\eta} (\varrho_\eta - \varrho_i + \varrho_i) = 1$$

Similarly, the values of y defined in (15.1)–(15.3) yield a solution of the system (12.1)–(12.h)–(13).

For a row $i \geq \eta$:

$$\varrho_\eta \frac{\varrho_h}{\varrho_\eta} \left[1 - \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \right]$$

$$+ \varrho_{\eta+1}\varrho_h \left(\frac{1}{\varrho_{\eta+1}} - \frac{1}{\varrho_\eta} \right) + \varrho_{\eta+2}\varrho_h \left(\frac{1}{\varrho_{\eta+2}} - \frac{1}{\varrho_{\eta+1}} \right) + ... + \varrho_i\varrho_h \left(\frac{1}{\varrho_i} - \frac{1}{\varrho_{i-1}} \right)$$

$$+(\varrho_i+\varrho_{i+1})\varrho_h \left(\frac{1}{\varrho_{i+1}} - \frac{1}{\varrho_i} \right) + (\varrho_i+\varrho_{i+2})\varrho_h \left(\frac{1}{\varrho_{i+2}} - \frac{1}{\varrho_{i+1}} \right) + ... + (\varrho_i+\varrho_h)\varrho_h \left(\frac{1}{\varrho_h} - \frac{1}{\varrho_{h-1}} \right) =$$

$$\varrho_h \left[1 + \varrho_i \left(\frac{1}{\varrho_h} - \frac{1}{\varrho_i} \right) \right] = \varrho_i.$$

For a row $i < \eta$:

$$(\varrho_i + \varrho_\eta)\frac{\varrho_h}{\varrho_\eta} \left[1 - \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \right]$$

$$+(\varrho_i+\varrho_{\eta+1})\varrho_h \left(\frac{1}{\varrho_{\eta+1}} - \frac{1}{\varrho_\eta} \right) + (\varrho_i+\varrho_{\eta+2})\varrho_h \left(\frac{1}{\varrho_{\eta+2}} - \frac{1}{\varrho_{\eta+1}} \right) + ... + (\varrho_i+\varrho_h)\varrho_h \left(\frac{1}{\varrho_h} - \frac{1}{\varrho_{h-1}} \right) =$$

$$\varrho_h \Bigg\langle \varrho_i \Bigg\{ \frac{1}{\varrho_{\eta+1}} - \frac{1}{\varrho_\eta} + \frac{1}{\varrho_{\eta+2}} - \frac{1}{\varrho_{\eta+1}} + ... + \frac{1}{\varrho_h} - \frac{1}{\varrho_{h-1}}$$

$$+ \frac{1}{\varrho_\eta} \left[1 - \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \right] \Bigg\}$$

$$+1-\frac{\varrho_{\eta+1}}{\varrho_\eta}+1-\frac{\varrho_{\eta+2}}{\varrho_{\eta+1}}+...+1-\frac{\varrho_h}{\varrho_{h-1}}+1 - \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \Bigg\rangle =$$

$$\varrho_h \left\{ 1 + \varrho_i \left[\frac{1}{\varrho_h} - \frac{1}{\varrho_\eta} \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \right] \right\}$$

$$= \varrho_h + \varrho_i \left[1 - \frac{\varrho_h}{\varrho_\eta} \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \right].$$

We need to show that (12.i) holds, *i.e.*:

$$\varrho_h + \varrho_i \left[-\frac{\varrho_h}{\varrho_\eta} \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \right] \leq 0.$$

Rearranging:

$$\frac{\varrho_\eta}{\varrho_i} \leq 1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}}$$

The above inequality holds for $i = \eta - 1$ since, by definition of η:

$$1 < 1 - \frac{\varrho_\eta}{\varrho_{\eta-1}} + 1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}}.$$

For $i < \eta - 1$, there suffices to note that:

$$\frac{\varrho_\eta}{\varrho_i} \leq \frac{\varrho_\eta}{\varrho_{\eta-1}}.$$

To complete the proof, the objective function can be calculated, using x:

$$\frac{1}{\varrho_\eta} \left\{ \varrho_h \left[1 - \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \right] \right.$$

$$+ \varrho_\eta \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} \right) + \varrho_{\eta+1} \left(1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} \right) + ... + \varrho_{h-1} \left(1 - \frac{\varrho_h}{\varrho_{h-1}} \right) \right\} =$$

$$\frac{1}{\varrho_\eta} \left[\varrho_h - \varrho_h \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right) + \varrho_\eta - \varrho_h \right] =$$

$$1 - \frac{\varrho_h}{\varrho_\eta} \left(1 - \frac{\varrho_{\eta+1}}{\varrho_\eta} + 1 - \frac{\varrho_{\eta+2}}{\varrho_{\eta+1}} + ... + 1 - \frac{\varrho_h}{\varrho_{h-1}} \right).$$

APPENDIX 2 - END OF PROOF OF PROPOSITION 2

Let $\sigma_i = \frac{\rho_{j_{i+1}}}{\rho_{j_i}}$ for $i = \kappa, ..., K-1$. With this expression, consider the relationship between the right-hand side of (7) and its parameters, *i.e.* the function E defined by $[0,1]^{K-\kappa} : (\sigma_\kappa, ..., \sigma_{K-1}) \hookrightarrow \sigma_\kappa \times ... \times \sigma_{K-1} \times (1 - \sigma_\kappa + ... + 1 - \sigma_{K-1})$. Note that the function E does not depend on κ or K per se, but on the difference $\acute{K} = K - \kappa$; thus, the notation $E(\sigma_1, ..., \sigma_{\acute{K}})$ will be sufficiently unambiguous. The proof is by induction.

When $\acute{K} = 1$, the apex of the parabolic function $\max_{\sigma_1 \in [0,1]} E(\sigma_1) = \left(\frac{1}{2} \right)^2$ for $\sigma_1 = \frac{1}{2}$. Assume now that Proposition 2 holds for $\acute{K} = k - 1$ (for any given value $k \geq 2$) *i.e.* $\left(\frac{k-1}{k} \right)^k = \max_{(\sigma_1, ..., \sigma_{k-1}) \in [0,1]^{k-1}} E(\sigma_1, ..., \sigma_{k-1})$; by proving that $\left(\frac{k}{k+1} \right)^{k+1} = \max_{(\sigma_1, ..., \sigma_k) \in [0,1]^k} E(\sigma_1, ..., \sigma_k)$, the latter property will in fact have been shown to hold for any value $\acute{K} \in \mathbb{N}$.

First, note that $E(\sigma_1, ..., 0, ..., \sigma_k) = 0$ and $E(\sigma_1, ..., \sigma_{i-1}, 1, \sigma_{i+1}, ..., \sigma_k) = E(\sigma_1, ..., \sigma_{i-1}, \sigma_{i+1}, ..., \sigma_k) \leq \left(\frac{k-1}{k} \right)^k < \left(\frac{k}{k+1} \right)^{k+1}$ (again, the last inequality is mentioned in [4]; the previous one comes from the recurrence hypothesis). Thus $\max_{\sigma_1, ..., \sigma_k \in [0,1]^k} E(\sigma_1, ..., \sigma_k) = \max_{\sigma_1, ..., \sigma_k \in]0,1[^k} E(\sigma_1, ..., \sigma_k)$: on the latter open set, the apex of E is characterized by the conditions $\frac{\partial E}{\partial \sigma_i} = \sigma_1 \times ... \times \sigma_k \times \sigma_i^{-1} (1 - \sigma_1 + ... + 1 - \sigma_k - \sigma_i) = 0$ for $i = 1, ..., k$. This system has a unique solution $\sigma_i = \frac{k}{k+1}$ for $i = 1, ..., k$, yielding $\left(\frac{k}{k+1} \right)^{k+1} = \max_{(\sigma_1, ..., \sigma_k) \in [0,1]^k} E(\sigma_1, ..., \sigma_k)$ as sought.

Appendix 3 - Extensive Proof of Corollary 2

First note that for consecutive integer coefficients, Bound (16) can be written:

$$\frac{z - z_G}{z - z_R} \le \frac{\rho_{j_H}}{\rho_{j_\hbar}} \left(\frac{1}{\rho_{j_\hbar}} + \ldots + \frac{1}{\rho_{j_{H-1}}} \right) \tag{19}$$

This proof consists of showing that the right hand side of Bound (18) is the largest among a family of bounds (16). Without any relation to the indices η or h used in Section 5 or Appendix 1, let $\eta = \rho_{j_\hbar}, h = \rho_{j_H}$, therefore $\rho_{j_{H-1}} = h + 1$ and the right hand-side of Bound (19) can be re-written: $\frac{h}{\eta} \left(\frac{1}{\eta} + \ldots + \frac{1}{h+1} \right)$. Lemma 2 shows that η is a (strictly) increasing function of $h < \delta$.

Lemma 2. If $\frac{1}{\eta} + \ldots + \frac{1}{h+1} \le 1$ then $\frac{1}{\eta+1} + \ldots + \frac{1}{h+2} < 1$.

Proof. $\frac{1}{\eta+1} - \frac{1}{h+1} < 0$ and thus $\frac{1}{\eta+1} + \ldots + \frac{1}{h+2} = \frac{1}{\eta+1} - \frac{1}{h+1} + \frac{1}{\eta} + \ldots + \frac{1}{h+1} < \frac{1}{\eta} + \ldots + \frac{1}{h+1} \le 1$.

$\square$

However, η can be such that:

$$\frac{1}{\eta + 1 + p} + \ldots + \frac{1}{h + 2} \le 1 \quad \text{for} \quad p = 1 \quad \text{or} \quad p = 2 \quad \text{or} \quad \ldots \quad \text{or} \quad p = d - \eta - 1. \tag{20}$$

Lemmata 3 and 4 show that, when $h < \delta$ (*i.e*, when η increases strictly with h), the right-hand side of (19) is a non-decreasing function of h.

Lemma 3. If $1 < \frac{1}{\eta+1} + \ldots + \frac{1}{h+1}$, then $\frac{h}{\eta} \left(\frac{1}{\eta} + \ldots + \frac{1}{h+1} \right) < \frac{h+1}{\eta+1} \left(\frac{1}{\eta+1} + \ldots + \frac{1}{h+2} \right)$.

Proof. First note that $\eta > h$, *i.e.* $\eta - h > 0$.

$$\frac{h+1}{\eta+1} \left(\frac{1}{\eta+1} + \ldots + \frac{1}{h+2} \right) - \frac{h}{\eta} \left(\frac{1}{\eta} + \ldots + \frac{1}{h+1} \right) =$$

$$\frac{h}{\eta(\eta+1)} - \frac{1}{\eta+1} + \left(\frac{h+1}{\eta+1} - \frac{h}{\eta} \right) \left(\frac{1}{\eta+1} + \ldots + \frac{1}{h+1} \right) =$$

$$\frac{h-\eta}{\eta(\eta+1)} + \frac{\eta-h}{\eta(\eta+1)} \left(\frac{1}{\eta+1} + \ldots + \frac{1}{h+1} \right) =$$

$$\frac{h-\eta}{\eta(\eta+1)} \left[1 - \left(\frac{1}{\eta+1} + \ldots + \frac{1}{h+1} \right) \right] > 0. \qquad \square$$

However, η considered in Lemma 3 may satisfy Relationship (20), a case addressed by the following

Lemma 4. If $\frac{1}{\eta+q} + \ldots + \frac{1}{h+2} \le 1$
then $\frac{h+1}{\eta+q} \left(\frac{1}{\eta+q} + \ldots + \frac{1}{h+2} \right) \le \frac{h+1}{\eta+q+1} \left(\frac{1}{\eta+q+1} + \ldots + \frac{1}{h+2} \right)$.

Proof. $\frac{h+1}{\eta+q+1}\left(\frac{1}{\eta+q+1}+\ldots+\frac{1}{h+2}\right)-\frac{h+1}{\eta+q}\left(\frac{1}{\eta+q}+\ldots+\frac{1}{h+2}\right)$

$= \frac{h+1}{(\eta+q)(\eta+q+1)}+\left(\frac{h+1}{\eta+q+1}-\frac{h+1}{\eta+q}\right)\left(\frac{1}{\eta+q}+\ldots+\frac{1}{h+2}\right)$

$= \frac{h+1}{(\eta+q)(\eta+q+1)}-\frac{h+1}{(\eta+q+1)(\eta+q)}\left(\frac{1}{\eta+q}+\ldots+\frac{1}{h+2}\right)$

$= (h+1)\frac{1-\left(\frac{1}{\eta+q}+\ldots+\frac{1}{h+2}\right)}{(\eta+q)(\eta+q+1)}\geq 0. \qquad \square$

Chaining Lemma 3 with the preceding result of Relationship (20) for $q = 1,...,p$, get:

$$\frac{h}{\eta}\left(\frac{1}{\eta}+\ldots+\frac{1}{h+1}\right) < \frac{h+1}{\eta+p}\left(\frac{1}{\eta+p}+\ldots+\frac{1}{h+2}\right).$$

Lemma 5 shows that, when $h > \delta$, *i.e.*, $\eta = d$, Bound (19) is a decreasing function of h.

Lemma 5. If $\frac{1}{d}+\ldots+\frac{1}{h+1}\ \ \leq 1$, then $\frac{h}{d}\left(\frac{1}{d}+\ldots+\frac{1}{h+1}\right)\geq\frac{h+1}{d}\left(\frac{1}{d}+\ldots+\frac{1}{h+2}\right)$

Proof. $\frac{h}{d}\left(\frac{1}{d}+\ldots+\frac{1}{h+1}\right)-\frac{h+1}{d}\left(\frac{1}{d}+\ldots+\frac{1}{h+2}\right) = \frac{1-\left(\frac{1}{d}+\ldots+\frac{1}{h+1}\right)}{d}\geq 0.$

Note that for $h \geq 0$, the inequalities of Lemma 5 hold strictly. $\qquad \square$

APPENDIX 4: ACTUAL VALUE OF $\frac{z-z_G}{z-z_R}$ FOR SOME REAL-WORLD DATA SETS

The same experimental suite is adopted to determine actual values of $\frac{z-z_G}{z-z_R}$ to be compared with Bound (16). The experimentation was limited by the relative difficulty of finding optimal solutions for Problem (1)–(4). Therefore the sample sizes (*i.e.*, the number of instances calculated for each case) were often decreased, as indicated at the top of each table when it is different from the corresponding table of Section 6. The number n of nodes was often limited to 30 (as also indicated in the tables) and therefore little sensitivity could be applied to finding out its impact on the performance of the Greedy solution. Quite generally, the ratio $\frac{z-z_G}{z-z_R}$ decreases with the size of K and experimentation is largely limited to $K = 8$. The import of the experiment itself is also limited by the relative lack of relationship between Ratio $\frac{z-z_G}{z-z_R}$ and Bound (16). One must further bear in mind that most practitioners are interested in the performance of the Greedy Heuristic with respect to the optimal value of Problem (1)–(5), not that of its linear relaxation, however close that latter two may be [1]. Significantly, the ratio $\frac{z-z_G}{z-z_R}$ is very small, compared with Bound (16), for the real-world data sets of [17,18] displayed in Tables 9–11, in comparison with Table 2. Only a few values are presented, as they stay systematically away from those of Bound (16).

TABLE 9. Percent value of $\frac{z-z_G}{z-z_R}$ for a data set in [17] with $n = 33$.

K	2	3	4	6	10	31
Percent	3	0	0	1	1	0

TABLE 10. Percent value of $\frac{z-z_G}{z-z_R}$ for a data set in [17] with $n = 57$.

K	3	4	5	6	29	55
Percent	1	0	0	0	0	0

TABLE 11. Percent value of $\frac{z-z_G}{z-z_R}$ for a data set in [18] with $n = 100$.

K	4	7	16
Percent	1	1	1

TABLE 12. Percent values of $\frac{z-z_G}{z-z_R}$ for complete Euclidian graphs with $n = 30$; sample size: 10.

p \ K	2	3	4	5	6	7	8
2	2	3	2	2	1	1	1
5	2	3	2	2	1	1	1

The ratio $\frac{z-z_G}{z-z_R}$ is similarly quite smaller than that of Bound (16) for complete Euclidian graphs. In Table 12 as in Table 3, it is not sensitive to the length of the square in which the nodes lie. Also as in Table 3, it decreases as K increases, a behavior displayed by all the following tables of the Appendix and largely explained (following Prop. 1) by the optimality of the Greedy heuristic when the number of locations selected does not reach K.

For trees with unit revenues shown in Table 13, the magnitude of the ratio $\frac{z-z_G}{z-z_R}$ becomes so small that it falls out the domain of magnitude of Bound (16). We delay a (limited) analysis of the influence of size to Table 14 because increasing size would not bring the ratio $\frac{z-z_G}{z-z_R}$ within distance of Bound (16).

TABLE 13. Permil value of $\frac{z-z_G}{z-z_R}$ for trees with unit revenues; sample size: 3.

n \ K	2	3	4	5	6	7	8
30	0	6	2	6	5	2	1

Furthermore, the ratio $\frac{z-z_G}{z-z_R}$ decreases markedly as the density of the graphs increases (following the experiment carried out in Tab. 6) and therefore it is not reported here. A possible explanation lies with the caution expressed at the end of the presentation of Table 6. With increased size, the distance between nodes becomes very small, with possibly little room for the Heuristic to err. The influence of the size of the graph is tested in Table 14, but because of experimental

limitations, the small range of variation of n does not permit to assert that it has little influence on the ratio $\frac{z-z_G}{z-z_R}$.

TABLE 14. Percent value of $\frac{z-z_G}{z-z_R}$ for Euclidian trees; sample size: 3.

n \ K	2	3	4	5	6	7	8
10	3	0	0	0	0	0	0
20	1	1	0	0	0	0	0
30	0	1	1	1	0	0	0
40	0	1	1	1	0	0	0

As the density of the tree increases in Table 15, following the protocol adopted in Table 7, the ratio $\frac{z-z_G}{z-z_R}$ increases until it reaches values comparable with those of Table 12 (the resemblance between the bottom of Tables 7 and 3 is less conspicuous).

TABLE 15. Percent value of $\frac{z-z_G}{z-z_R}$ for sparse Euclidian graphs with $n = 30$; sample size: 3.

q \ K	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1
2	2	2	1	1	1	1	1
3	1	1	0	0	0	0	0
4	1	1	1	1	1	1	1
5	0	0	0	0	0	0	0
6	2	1	1	1	0	0	0
7	0	1	1	0	0	0	1
8	1	1	1	1	1	1	1
9	1	0	1	0	0	0	0
10	3	2	2	1	1	1	1
11	2	2	2	1	1	1	1
12	2	2	2	2	2	1	1
13	2	2	2	2	2	1	1

Finally, the variation of the ratio $\frac{z-z_G}{z-z_R}$ in Table 16 is not in contrast with that of preceding tables, contrarily to Table 8. It is not affected by the magnitude of the random revenues and therefore only one range of magnitude is displayed.

TABLE 16. Percent values of $\frac{z-z_G}{z-z_R}$ for random revenues with $n = 30$; sample size: 10.

r \ K	2	3	4	5	6	7	8
2	2	2	1	1	1	1	0

REFERENCES

[1] S. Ahn, C. Cooper, G. Cornuéjols and A. Frieze, Probabilistic Analysis of a Relaxation for the k-Median Problem. *Math. Oper. Res.* **13** (1988) 1–31.

[2] V. Chvátal, A Greedy Heuristic for the Set-Covering Problem. *Math. Oper. Res.* **4** (1979) 233–235.

[3] G. Cornuéjols, M.L. Fisher and G.L. Nemhauser, On the Uncapacitated Location Problem. *Ann. Discrete Math.* **1** (1977) 163–178.

[4] G. Cornuéjols, 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** (1977) 789–810.

[5] G. Cornuéjols, M.L. Fisher and G.L. Nemhauser Note on Location of Bank Accounts to Optimize Float: An Analytic Study of Exact and Approximate Algorithms. *Management Science* **25** (1979) 808–809.

[6] G. Cornuéjols and J.-M. Thizy, *Location Problems, Set Covering Problems and the Greedy Algorithm*, Working paper 25-80-81, Graduate School of Industrial Administration, Carnegie-Mellon University (April 1981), also available as Working Paper 02-51, School of Management, University of Ottawa (2002).

[7] G. Cornuéjols and J.-M. Thizy, *New Results on the Greedy Algorithm for Plant Location and Set Covering Problems*, presented at the CORS-TIMS-ORSA Joint National Meeting, Toronto, Ontario, May 4, 1981.

[8] G. Cornuéjols and J.-M. Thizy, A Primal Approach to the Simple Plant Location Problem. *SIAM J. Algebraic Discrete Methods* **3** (1982) 504–510.

[9] G. Cornuéjols, G.L. Nemhauser and L.A. Wolsey, The Uncapacitated Facility Location Problem, in: *Discrete Location Theory*, P.B. Mirchandani and R.M. Francis Eds., John Wiley and Sons, New–York (1990) 119–171.

[10] A.M. El-Shaieb, A New Algorithm for Locating Sources Among Destinations. *Management Science* **20** (1973) 221–231.

[11] D. Erlenkotter, A Dual-based Procedure for Uncapacitated Facility Location. *Oper. Res.* **26** (1978) 992–1009.

[12] S. Even, *Graph Algorithms*. Computer Science Press, Potomac, Maryland (1979).

[13] M.L. Fisher, G.L. Nemhauser and L.A. Wolsey, An Analysis of Approximations for Maximizing Submodular Set Functions-II, *Mathematical Programming Study* **8** (1978) 73–87.

[14] M.L. Fisher and L.A. Wolsey, On the Greedy Heuristic for Covering and Packing Problems, *SIAM J. Algebraic Discrete Methods* **3** (1982) 584–591.

[15] M. Guignard and K. Spielberg, Algorithms for exploiting the Structure of the Simple Plant Location Problem. *Ann. Discrete Math.* **1** (1977) 247–271.

[16] D.S. Johnson, Approximation Algorithms for Combinatorial Problems. *J. Comput. System Sci.* **9** (1974) 256-298.

[17] R.L. Karg and G.L. Thompson, A Heuristic Approach to Solving Traveling Salesman problems, *Management Sci.* **10** (1964) 225–248.

[18] P. Krolak, W. Felts and G. Marble, A Man-Machine Approach toward Solving the Traveling Salesman Problem. *Communications of the Association for Computing Machinery* **14** (1971), 327–334.

[19] A.A. Kuehn and M.J. Hamburger, A Heuristic Program for Locating Warehouses, *Management Sci.* **9** (1963) 643–666.

[20] G.L. Nemhauser, L.A. Wolsey and M.L. Fisher, An Analysis of Approximations for Maximizing Submodular Set Functions-I. *Math. Program.* **14** (1978) 265–294.

[21] L. Schrage, Implicit Representation of Variable Upper Bounds in Linear Programming, *Math. Program. Stud.* **4** (1975) 118–132.

[22] H.P. Simão and J.-M. Thizy, A Simplex Algorithm for the Canonical Representation of the Uncapacitated Plant Location Problem. *Oper. Res. Lett.* **8** (1989) 279–286.

[23] J.-M. Thizy, *Location Problems: Properties and Algorithms*, Ph.D. thesis, Graduate School of Industrial Administration, Carnegie Mellon University (1981).

[24] J.-M. Thizy, *Worst-case analysis of the uncapacitated facility location problem with a budget constraint*, Working Paper, School of Management, University of Ottawa, to appear.

[25] L.A. Wolsey, An analysis of the Greedy Algorithm for the Submodular Set Covering Problem'. *Combinatorica* **2** (1982) 417–425.