Optimizing Facility Location and Design

Robert Aboolian[(a)]*, **Oded Berman**[(b)]
Dmitry Krass[(b)]

[(a)]College of Business Administration, California State University San Marcos
San Marcos, California, 92096, USA

[(b)]Joseph L. Rotman School of Management, University of Toronto
105 St. George Street, Toronto, Ontario, Canada M5S 3E6

April 9, 2012

Abstract

In this paper we develop a novel methodology to simultaneously optimize locations and designs for a set of new facilities facing competition from some pre-existing facilities. Known as the Competitive Facility Location and Design Problem (GFLDP), this model was previously only solvable when a limited number of design scenarios was pre-specified. Our methodology removes this limitation and allows for solving of much more realistic models. The results are illustrated with a small case study.

*Corresponding author

1 Introduction

OUTLINE: Relationship to GFLDP. Limitation of the general GFLDP with design scenarios: where do scenarios come from? Goal of current paper is to provide a complete solution to the GFLDP with unlimited number of potential designs.

We start by focusing on the single-facility GFLDP, developing an efficient polynomial-time algorithm for this case. This is accomplished by showing that the problem can be decomposed into two sub-problems: the Single-Facility Design Problem (SFDP), which can be solved efficiently, and a simple search problem. Applying parametric analysis to the optimal solution for (SDFP) we show, under relatively mild assumptions, that the facility attractiveness is a concave function of the design budget; this function is derived in closed form.

We next apply this methodology to the multiple-facility GFLDP. The basic problem is shown to be non-linear in both the objective and the constraints. However, most of the non-linearities can be expressed with concave functions. By building on the TLA-method developed earlier, which allows for an efficient linear approximation of a concave function to the specified degree of accuracy, we develop an "iterated TLA" approach which allows us to solve the CDFLP by solving a single linear integer program.

We analyze the efficiency of this approach through a set of numerical experiments and illustrate our model through a small case study.

2 Definitions and Preliminaries

We assume that customer demand is concentrated in a discrete subset N of size $n = |N|$ of metric space $\mathcal{P}$ equipped with a distance function $d_{ij},\ \ i,j \in N$ (if $\mathcal{P}$ represent a network with node set N, the function d_{ij} can be chosen to be the shortest path distance). Each of the n points represents a "market", i.e., a set of customers who are homogeneous with respect to their facility preferences and expenditure decisions; in view of assumed homogeneity we will often refer to each $i \in N$ as "customer i". The maximum potential demand for $i \in N$ is given by w_i; this can be usefully viewed as the available "expenditure budget" for customer

i, who, depending on the attractiveness and convenience of service offered by the facilities, may choose to spend all, part, or none of it.

The set of potential facility locations P is assumed to be discrete; without loss of generality, we assume $P \subset N$ (note that dummy customer nodes with $w_i = 0$ can always be added to N). There may be pre-existing competitive facilities on the network located in set C with $P \cap N^C = \emptyset$.

As discussed in the introduction, our goal is to develop a modeling framework that is flexible enough to incorporate a wide variety of facility location and demand models, including most of the typical models in location literature. We first describe the basic components of our modeling structure, and then provide more details for each component below.

- The decision-maker (DM) makes two basic sets of decisions: "where" and "what". More specifically, the first decision is to select a set $S \subseteq P$ of locations for the new facilities, which includes deciding on $|S|$ - the number of new facilities to be opened. Then, for each $j \in S$, the DM must determine the "design" of facility j by specifying the values of $|K|$ design characteristics, yielding the attractiveness value of A_j. The mechanism underlying A_j is described later. Note that to simplify the notation (and the task of estimating the model parameters), we will assume that the perception of attractiveness A_j of facility j is the same for all customers $i \in N$, though some extensions to customer-specific forms will be discussed later. The locations of opened facilities and their design (as well as the location and design of competitive facilities) is viewed as "offered service" by the customers.

- Each customer $i \in N$ evaluates the individual attractiveness (utility) of each facility $j \in S$:

$$u_{ij} = A_j g(d_{ij}), \text{where } g(d) \in [0, 1], g(0) = 1 \text{ is non-increasing and convex in } d \quad (1)$$

  Note that since u_{ij} which depends on both the proximity d_{ij} of faclity j and its perceived attractiveness, the utilities different customers derive from facility j are different, even though the perceived attractiveness A_j is the same for all customers. The function $g(d) \geq 0$ represents distance sensitivity of customers; it is assumed to be non-decreasing with $g(0) = 1$.

- Next each customer computes the total value (utility) of all new facilities $U_i(S), i \in N$. We consider two alternative forms of this total utility function:

$$U_i^A(S) = \sum_{i \in S} u_{ij} \tag{2}$$

$$U_i^M(S) = [\max_{j \in S} u_{ij}] * I\{\max_{j \in S} u_{ij} \geq \max_{j \in C} u_{ij}\} \tag{3}$$

  Thus, under $U_i^A(S)$ specification, the customer i simply adds utilities derived from all facilities in S. As we will see later, this specification is required to represent location models with proportional or "gravity-type" assignments, where customers patronize more than one facility and facilities compete for market share of customer's expenditures. The second mechanism, $U_i^M(S)$ is needed to represent all-or-nothing assignments, where customer only patronizes the utility-maximizing facility. Note that if customers' utility is maximized by a facility belonging to the competitor (i.e. $j \in C$) then $U_i^M(S) = 0$, otherwise it is set to the maximum utility derived from $j \in S$. Of course other utility aggregation mechanisms can be specified as well, but as we will see below, these two mechanisms provide sufficient flexibility to represent many previously-described location models. When either mechanism can be used, we will simply refer to $U_i(S)$ as the total utility.

- The final decision made by the customer $i \in N$ is how much of the available budget to spend on the service offered by the facilities in S. We use $V(U_i(S))$ to represent the *proportion* of the available expenditure demand that customer $i \in N$ spends on service offered by facilities in S, where we assume that $V(U)) \in [0, 1], V(0) = 1, \lim_{U\to\infty} V(U) = 1$ and $V(U)$ is concave, non-decreasing, and twice-differentiable function in U. The total expenditure for $i \in N$ is given by $w_i V(U_i(S)$. Note that even though the form of the demand function $V(U)$ is assumed to be the same for all customers, no substantial difficulties arise from making this form customer-specific.

- Finally, the model is completed by specifying the objective of the DM. Letting $C(U(S), S)$ represent the cost of locating new facilities in set S and ensuring utility vector $U(S) = [U_i(S), \ldots, U_n(S)]$ for customers in N (we will specify the functional form of $C(U, S)$ below), the GDFLP can be stated in compact form as follows:

$$\max Z(S) = \{\sum_{i \in N} w_i V(U_i(S)) \mid S \subset P, |S| \leq m, C(U(S), S) \leq b\}, \tag{4}$$

  where m represents the maximum number of facilities that can be located, and b represents the total budget available to the DM. An alternative formulation, where DM seeks to maximize net revenue and the budget constraint is dropped, will be discussed in Section 6.

To complete the model description we next specify how A_j is related to design decisions and the associated costs.

Facility Design

Following [2], we assume that each potential site j possesses a "base" attractiveness level of $\alpha_j \geq 0$ which can be interpreted as the attractiveness of the basic ("bare-bones") facility that could feasibly be located at j - i.e., that would satisfy all required regulations and municipal codes with respect to safety, parking, etc. at the minimal level. Note that α_j incorporates site-specific characteristics such as visibility, ease of access, etc. We also assume that there is set K of design characteristics with respect to which the facility can be further improved, with the decision variables $Y_{jk} \geq 0$ representing the *improvement* of the basic design along characteristic $k \in K$, with the value $Y_{jk} = 0$ if facility at site $j \in S$ is not improved with respect to characteristic k. The resulting attractiveness A_j of a facility at $j \in P$ is assumed to be a log-linear function:

$$A_j = \alpha_j \prod_{k \in K} (1 + Y_{jk})^{\theta_k} \tag{5}$$

(throughout this paper we adopt the convention that the product evaluated over an empty set equals to 1, thus $A_j = \alpha_j$ when $K = \emptyset$). As discussed in [2], this form reflects the most common approaches to estimating the impact of characteristics of goods and services using Multi-Competitive Interaction (MCI) and related models (see e.g., Nakanishi and Cooper [14]); the methodology for estimating the values of $\theta_k, k \in K$ is well-developed.

We will typically assume that $Y_{jk} \in [0, y_k^{max}]$, i.e., the value of each design characteristic can be adjusted continuously within this interval, even though for some qualitative characteristics only discrete values may make sense; the extension to incorporate qualitative characteristics will be discussed in Section **??** below. We make the following assumption regarding the parameters $\theta_k, k \in K$:

Assumption 1: For all $k \in K$, $\theta_k \leq 1$.

This assumption ensures that A_j is concave with respect to every design variable $Y_{jk},\ k \in K$, implying that the marginal attractiveness is decreasing as the design improvements are made; this is a very common assumption is economics literature. For some of our results it will be necessary to make a stronger assumption:

Assumption 2: $\sum_{k \in K} \theta_k \leq 1$.

As will be shown below, Assumption 2 ensures that A_j is concave with respect to the overall expenditure on improvements for facility j; we believe this assumption should also be reasonable in most cases. We will indicate which of our results below require the stronger Assumption 2 instead of Assumption 1.

We note that it may be natural to regard travel distance as just another characteristic of a facility, which results in the following specification of the term $g(d)$ in (1):

$$g(d) = (1 + d)^{-\beta}, \tag{6}$$

where $\beta > 0$ is a distance elasticity parameter playing the same role as $\theta_k, k \in K$ above. This leads to log-linear form for the individual attractiveness terms $u_{ij}, i \in N, j \in S$. While this form simplifies practical task of estimating model parameters (since the MCI methodology can be used), the specific form of $g(d)$ is not required in our methodology as long as the previously-stated properties (non-decreasing, $g(0) = 1$) are satisfied.

To complete the formulation we define the facility costs as follows. For a potential location $j \in P$, the parameter $f_j \geq 0$ represents the fixed cost of locating a facility with basic design at j, while c_{jk} is the variable cost of improving the design characteristic k by one unit (we assume that both fixed and variable costs are properly annualized). Thus, the cost of locating a facility with design attributes $Y_{j1}, \ldots, Y_{jK}$ is given by

$$B_j = f_j + \sum_{k=1}^{K} c_{jk} Y_{jk}, \tag{7}$$

where B_j (a decision variable) is the total allocated budget for a new facility at $j \in P$.

This expression assumes linear cost structure with respect to design characteristics. While this is likely a simplification of the actual costs involved in design decisions, the linearity enables us to develop efficient solution approaches below; we will discuss possible generalizations in Section 6. In view of the previous expression, the overall costs of locating facilities in set S to induce the customer utility vector $U(S)$ are given by

$$C(U(S); S) = \sum_{j \in S} B_j,$$

and the budget constraint in GFLDP formulation (4) can be rewritten as $\sum_{j \in S} B_j \leq b$.

This completes the formulation of GFLDP. Before developing solution strategies for this model, we first demonstrate the flexibility of the framework outlined above by showing how it generalizes a number of previously analyzed location problems.

2.1 Location Models with Proportional Allocation: GFLDP Representation

In this class of models each customer is assumed to divide their spending between several (potentially all) open facilities in proportion to the utility derived from each facility. These models, which have a number of names (gravity-type, Huff-type, market share games, competitive interaction models, multinomial logit models) date back to the work of Huff [13]; see Gosh *et al* [10] for a good introduction and Berman *et al* [4] for a more recent overview. The basic model of this type assumes inelastic demand, (i.e. each customer $i \in N$ spends the maximum amount w_i) and no flexibility in facility design (i.e., the set of changeable design characteristics $K = \emptyset$). The utility customer $i \in N$ derives from each facility $j \in S \cup N^C$ is given by

$$u_{ij} = \alpha_j g(d_{ij}),$$

where $g(d)$ is some decreasing function of distance (note that for some popular forms, e.g. $g(d) = d^{-2}$, which was used in Huff's original paper [13] and many subsequent publications, one must make sure that some minimal distance $0 < d_\epsilon < d_{ij}$ exists for all $i \in N, j \in S$).

Once a location set is defined, the market shares are computed via

$$MS_{ij} = \frac{u_{ij}}{\sum_{k \in S \cup N^C} u_{ik}}, i \in N, j \in S \cup N^C \tag{8}$$

and the problem is to find a location set $S, |S| \leq m$ to maximize

$$Z_{PA}(S) = \sum_{i \in N} \sum_{j \in S} w_i MS_{ij}. \tag{9}$$

In words, each facility (which includes both, the new facilities in S and competitive facilities in N^C) receives the share MS_{ij} of customers demand w_i that is proportional to its utility u_{ij} for customer $i \in N$; the goal is to intercept as much demand at the new facilities as possible. We also observe that $\sum_{j \in S \cup N^C} MS_{ij} = 1$ and thus the problem loses meaning when $N^C = \emptyset$, as $Z_{PA}(S) = \sum_{i \in N} w_i$ irrespective of S. To recast this problem into our framework we observe that since there are no design characteristics to optimize, $A_j = \alpha_j$, and thus u_{ij} is already in the form given by (1) above. We use the additive form (2) to define $U_i = \sum_{i \in S} u_{ij}$ - the total utility customer i derives from the new facilities in S and define the demand function

$$V(U_i(S)) = \frac{U_i}{U_i + U_i^C}, \tag{10}$$

where $U_i^C = \sum_{j \in N^C} u_{ij}$ is the total utility derived by $i \in N$ from the competitive facilities - a constant that can be pre-computed since N^C is assumed to be known. It is easy to see that $Z_{PA}(S) = \sum_{i \in N} w_i V(U_i(S))$ and that $V(U)$ is concave and non-decreasing in U. Thus, the basic proportional allocation model is easily representable in our GFLDP framework.

We also discuss two extensions of the basic model. The first one, described in Berman and Krass [5], introduced the elastic demand $w_i e_i(U_i) = e(\sum_{j \in S \cup N^C} u_{ij})$ where $e(U) \in [0,1]$ is a concave non-decreasing function of the total utility the customer derives from all facilities on the network, in place of w_i in the objective function Z_{PA} in (9). This makes the proportional allocation model meaningful even when $N^C = \emptyset$ since locating new facilities increases the overall utility and thus captures more of customer's potential expenditure w_i (which is now treated as the maximal expenditure of customer $i \in N$, same as in our GFLDP). This elastic demand extension is also easily representable as GFLDP by simply replacing U_i with $U_i e(U_i + U_i^C)$ in the numerator of $V(U_i(S))$ in (10) above.

The idea of flexible design was introduced in Aboolian *et al* [2] using essentially the same framework as in the current paper, with the key difference that only a finite number of pre-determined design scenarios was allowed. Thus, GFLDP can be seen as an extension where an infinite number of potential design scenarios is considered for each facility.

2.2 Full Capture Location Models: GFLDP Representation

The second major class of location models (in terms of customer-facility interaction) are the "full-capture" (also known as "all-or-nothing" models) where each customer $i \in N$ is assumed to patronize only the facility j^* that is utility-maximizing for them. This class of models date back to the work of Revelle [15] and Hakimi [11]. We start with the MAXCAP model of [15], which generalizes $m-$median and related model. It is assumed that for each $j \in P \cup N^C$, a customer at $i \in N$ derives "benefit" r_{ij} when obtaining service from facility j, and the objective is to find $S, |S| \leq m$ which maximizes

$$Z_{FC}(S) = \sum_{i \in N} \sum_{j^* \in S} w_i r_{ij^*},$$

where $j^* = \arg\max_{j \in S \cup N^C} r_{ij}$ with ties typically broken in favor of the new facility (i.e., if the maximum benefit is achieved at both an existing and a new facility, then $j^* \in S$ is selected). To recast this into GFLDP form we can simply treat the benefits r_{ij} as reciprocals of distances $d_{ij} = 1/r_{ij}$ after assuming, without loss of generality, that r_{ij} values are scaled to be in $(0, 1]$. Next we set $K = \emptyset$, $\alpha_j = 1$ for all $j \in P \cup N^C$ and $g(d) = 1/d$. It can be seen that $u_{ij} = r_{ij}$ in this case. Selecting $U_i = U^M$ in (3) we see than $U_i = r_{ij^*}$ for all i and the with $V_i(U_i) = U_i$ we have $Z_{FC}(S) = \sum_{i \in S} w_i V(U_i(S))$ showing that MAXCAP is a special case of GFLDP.

We can extend the basic model by allowing for elastic demand: this is easily accomplished by re-defining $V_i(U)$ to be a concave non-decreasing function of U with range in $[0, 1]$. We can also add the design component by re-defining $r_{ij} = A_j(\mathbf{y})g(d_{ij})$, where A_j and $g(D)$ are given by (5), (1), respectively.

Having demonstrated the versatility of GFLDP framework we next address the solvability issues. While the (GFLDP) can be formulated as a non-linear integer program using the standard binary location variables X_j augmented with the design variables Y_{jk} for $j \in S, k \in K$, the practicality of this formulation is doubtful since the demand function $V_i(U_i(S))$ is a non-linear functions of the utility vector, which, in turn, is non-linear in the design variables $\mathbf{Y}$. Thus, direct solution approaches are unlikely to be successful. We thus develop an alternate approach by first focusing on the single-facility optimal design problem which arises once the decision to locate a facility at $j \in P$ has been made.

3 Single Facility Design Problem (SFDP)

In this section we first assume that we have already decided to locate a new facility at some location $j \in P$ and spend at most $B_j \leq b$ on this facility. The problem we address is how to determine the optimal design of this facility for the given budget B_j. We then perform parametric analysis with respect to B_j to identify the optimal design for any budget level. This problem may be of interest in its own right for the case where only one new facility is to be located: se set $b_j = b$ and solve SDFP for every potential $j \in P$. With respect to the original multi-facility setting, the methodology we develop for SDFP allows the decision-maker to reduce the dimensionality the one involving multiple design dimensions $y_{jk}, k = 1, \ldots, K$ to just one dimension, $B_j \in [0, b]$ for every $j \in P$. We will use this reduction in Section 4 to construct efficient algorithms for the multiple-location (GFLDP).

Suppose we are designing a new facility to be constructed at location j. To simplify the notation in this section, we will drop the subscript j (replacing f_j by f, etc.). Since the cost of the basic design is f, this implies that $B \geq f$ (otherwise no facility can be constructed at j) and the amount to be spent on design improvements is given by $B - f$. Note that the overall objective $Z(S)$ given by (4) above is clearly non-decreasing in the components u_i of the utility vector. Moreover, by (1), each utility component u_i is an increasing linear function of the attractiveness A. It follows that the objective function of (GFLDP) is maximized by maximizing A, leading to the following formulation for (SFDP):

$$\max A_j(B_j) \equiv A(B) = \alpha \prod_{k=1}^{K} (1 + y_k)^{\theta_k} \tag{11}$$

Subject to

$$\textstyle\sum_{k=1}^{K} c_k y_k \leq B - f \tag{12}$$

$$0 \leq y_k \leq y_k^{max}, \quad k \in K. \tag{13}$$

We will assume throughout this section that $0 < B - f =< \sum_{k \in K} c_k y_k^{max}$, since otherwise the problem is either infeasible, or admits the trivial optimal solution $y_k = y_k^{max}$ for all k.

3.1 Solving the SFDP for a Given B

The (SFDP) can be recognized as a non-linear knapsack problem. Moreover, by Assumption 1 we have $\theta_k \leq 1$ for all $k \in K$, implying that the objective function is concave. Taking logarithm of the objective, we see that (SFDP) is equivalent to a separable, concave knapsack problem. This class of problems is fairly well-solved; the complexity is known to be not much harder than that of the linear integer knapsack problem - we refer the reader to Bretthauer and Shetty [8] for a review.

We start with the following result, that develops the closed-form solution for the case where the upper and lower bounds in (13) are automatically satisfied by the unconstrained solution.

Lemma 1 *For $k \in K$, let*

$$\hat{y}_k(K,B) = \frac{\theta_k(B - f + \sum_{r\in K} c_r)}{c_k \sum_{r\in K} \theta_r} - 1. \tag{14}$$

If $0 \leq \hat{y}_k(K,B) \leq y_k^{max}$ for all $k \in K$, then $\hat{y}_k(K,B)$, $k \in K$ is an optimal solution to (SFDP).

Proof: For convenience, we make the variable substitution $\underline{y}_k = y_k + 1$, let $\underline{B} = B - f + \sum_{k\in K} c_k$, and drop the bounds (13). This leads to the following problem:

$$\max\{\prod_{k=1}^{K} \underline{y}_k^{\theta_k} | \sum_{k=1}^{K} c_k \underline{y}_k \leq \underline{B}\}. \tag{15}$$

Since this is a concave maximization problem with linear constraints, the first-order Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for optimality, provided the lower bounds on $\underline{y}_k$ are satisfied. This results in the following system of equations, where λ is the multiplier associated with the knapsack constraint:

$$\underline{y}_k = (\prod_{r\in K} \underline{y}_r^{\theta_r}) \frac{\theta_k}{\lambda c_k}, \tag{16}$$

and $\sum_{k\in K} c_k \underline{y}_k = \underline{B}$ (observe that since the upper bounds have been relaxed, the knapsack constraint must be tight at optimality), together with $\lambda \geq 0$. Substituting the first equation above into the second, we obtain

$$\lambda = (\prod_{r\in K} \underline{y}_r^{\theta_r}) \frac{\sum_{r\in K} \theta_r}{\underline{B}} \geq 0,$$

which, when substituted back into (16) yields $\underline{y}_k = \hat{y}_k(K,B) + 1$, $k \in K$. Therefore, $\hat{y}_k(K,B)$, $k \in K$ solves (SFDP), since the upper and lower bounds (13) are satisfied by hypothesis.

$\square$

We now present the following algorithm for (SFDP).

Algorithm 1

Step 0 Set $i = 0$, $K^i = K$, $U^i = L^i = \emptyset$, $B^i = B - f$.

Step 1 Set $i = i + 1$. Compute $\hat{y}_k(K^i, B^i)$, $k \in K$ using (14) and set $y_k^i = \hat{y}_k(K^i, B^i)$ for $k \in K^i$, $y_k^i = 0$ for $k \in L^i$ and $y^i = y_k^{max}$ for $k \in U^i$.

If $0 \leq y_k^i \leq y_k^{max}$ for $k \in K^i$, STOP and output current $y_k^i, k \in K$ as the optimal solution to (SFDP); else, proceed to Step 2.

Step 2 Let

$$\bar{s} = \sum_{k \in K^i} c_k(y_k^i - y_k^{max}) \cdot I\{y_k^i > y_k^{max}\}; \quad \underline{s} = \sum_{k \in K^i} c_k(-y_k^i) \cdot I\{y_k^i < 0\},$$

where $I\{\ \}$ is the indicator function. If $(\underline{s} \geq \bar{s})$ then set $L^{i+1} = L^i \bigcup\{k \in K^i | y_k^i < 0\}, B^{i+1} = B^i$

Else set $U^{i+1} = U^i \bigcup\{k \in K^i | y_k^i > y_k^{max}\}, B^{i+1} = B - \sum_{k \in U^{i+1}} y_k^{max} c_k$.

Set $K^{i+1} = K - U^{i+1} - L^{i+1}$, $i = i + 1$, and repeat Step 1.

At each step i the sets K^i, U^i, and L^i represent the variables that are "free", have been "pegged" to the upper bound y_k^{max}, and have been pegged to the lower bound 0, respectively. The budget B^i represents the available budget for the free variables, which is the original budget minus what has been used on the pegged variables in U^i (the variables in L^i are set to 0, and thus do not consume any budget). At iteration i we compute the unconstrained solution relaxation (14) over the "free" variables with the remaining budget. If this solution is feasible, it is optimal. Else, the total sum of violations of the upper and lower bounds is computed and, depending on which is larger, corresponding variables are pegged to their upper or lower bounds; the process continues until no violations remain. Since at each step of the algorithm, at least one new variables is pegged, the number of iterations is at most $|K|$ and the total time complexity is at most $O(|K|^2)$ (since in practical applications the number of design characteristics which can be varied is typically small, the algorithm will execute essentially instantaneously).

The correctness of the algorithm is established in the following result. Part (1) follows from Lemma 1 above and Theorem 4 in [7]; part (2) follows since the optimal solution must satisfy the KKT conditions.

Theorem 1 *Assume* $B < \sum_{k\in K} c_k y_k^{max}$. *Then:*

(1) Algorithm 1 computes an optimal solution $y_k, k \in K$ *to (SFDP) in* $O(|K|^2)$ *time.*

(2) Let L, U *be the final values of sets* L^i, U^i *produced by Algorithm 1. Then*

$$\hat{y}_k(K - L - U, B - f - \sum_{r\in U} c_r y_r^{max}) \begin{cases} < 0 & \text{if } k \in L \\ > y_k^{max} & \text{if } k \in U \\ = y_k \in [0, y_k^{max}] & \text{if } k \in K - L - U \end{cases} \tag{17}$$

It is interesting to observe from equations (14) how the available budget is allocated: the amount allocated to characteristic k depends on the relative attractiveness $\theta_k / \sum_{r\in K} \theta_r$ divided by the unit cost c_k, i.e., larger budget will be allocated to the characteristics that have higher relative attractiveness per dollar. Note also that this shows that in estimating θ_k parameters, only the relative values have the effect on the optimal design. We illustrate Algorithm 1 and the previous result with the following example.

Example 1. Consider an instance of (SFDP) with three design characteristics and the following data:

$$y^{max} = (1, 1, 1), \; c = (.5, 1, 2), \quad \text{and } \theta = (.5, .3, .1)$$

(where we provide components of the corresponding vectors). Suppose $f = 0$ and $B = .7$. In the first iteration of Step 1 of Algorithm 1 we compute $\hat{y}_k(K, B)$ with $K = \{1, 2, 3\}$, obtaining:

$$y^1 = \hat{y}(K, .7) = (3.667, .4, -.767).$$

Note that $\hat{y}_1 > y_k^{max}$, and $\hat{y}_3 < 0$, thus the current solution is not feasible and we compute

$$\bar{s} = c_1 * (\hat{y}_1 - y_1^{max}) = 1.333, \ \underline{s} = c_3 * (-\hat{y}_3) = 1.533.$$

Since $\underline{s} > \bar{s}$, we peg the last component to 0, getting $L^1 = \{3\}$, $U^1 = \emptyset, B^1 = B = .7$ and proceed to iteration 2. After two more iterations the algorithm terminates with the solution:

$$y = (1, .2, 0), \ L = \{3\}, \ U = \{1\},$$

indicating that the first component has been pegged at the upper bound and the third at the lower bound. Note that $K - L - U = \{2\}$, $B - \sum_{r \in U} c_r y_r^{max} = .7 - .5 = .2$, and $\hat{y}(\{2\}, .2) = (3, .2, -.8)$, verifying part 2 of Theorem 2 since the component in U exceeds the upper bound and the component in L is negative.

3.2 Parametric Analysis with Respect to Budget B

The previous results also allow us to perform parametric analysis of the SFDP with respect to the available budget B. This analysis will be used to extend the single-facility results to multi-facility settings in Section 4 below.

Note that, in view of (11, 14) and (17), the total attractiveness A can be viewed as a function $A(B)$ of the specified budget B:

$$A(B) = \alpha \prod_{k \in K-L-U} [(B - f + \sum_{r \in K-L-U} c_r - \sum_{r \in U} c_r y_r^{max}) \frac{\theta_k}{c_k \sum_{r \in K-L-U} \theta_r}]^{\theta_k} \prod_{k \in U} (y_k^{max} + 1)^{\theta_k}. \quad (18)$$

For fixed sets L, U, let

$$\gamma = f - \sum_{r \in K-L-U} c_r + \sum_{r \in U} c_r y_r^{max},$$

and

$$\delta = \prod_{k \in K-L-U} [\frac{\theta_k}{c_k \sum_{r \in K-L-U} \theta_r}]^{\theta_k} \prod_{k \in U} (y_k^{max} + 1)^{\theta_k}.$$

Then, (18) takes the form

$$A(B) = \alpha\delta(B-\gamma)^{\sum_{k\in K-L-U}\theta_k}, \tag{19}$$

where $\delta > 0$ and $\gamma < B$. This function is continuous and increasing in B and is concave in B if $\sum_{k\in K-L-U}\theta_k \leq 1$, i.e., when Assumption 2 holds. This also provides an interpretation of Assumption 2: it ensures that $A(B)$ is concave, i.e., that the marginal attractiveness is non-increasing per unit expenditure on design improvements (versus Assumption 1 which imposes a similar condition with respect to each individual design characteristic).

Observe, however, that the sets L, U are themselves affected by the value of B; thus the statements above holds only for values of B where sets L and U do not change. It is obvious from (14) and (17) that for any value of B we can find values $\underline{B}, \bar{B}$ such that sets L, U are invariant in the interval $B \in [\underline{B}, \bar{B}]$. It then follows from (19) that for $B \in [\underline{B}, \bar{B}]$, $A(B)$ is non-decreasing, and is concave in B if Assumption 2 holds. The breakpoints $\underline{B}, \bar{B}$ designate budget levels where the membership of sets L, U changes.

Determining the breakpoints is computationally easy. Consider first the upper breakpoint $\bar{B}$. Since $\hat{y}_k(K, B)$ is linear in B, if $\hat{y}_k(K-L-U, B-\sum_{j\in U} c_j y_j^{max}) > y_k^{max}$ for the current level of B, it will continue to be so as B is increased. Thus, as B is increased, any $k \in U$ will remain in set U. Similarly, any $k \notin L$ will remain outside of L (since the corresponding component of $\hat{y}_k$ will remain positive). Thus, $\bar{B}$ is reached when either some element leaves L or some element enters U. Consider the former case first. Suppose $k \in L$, implying that $\hat{y}_k(K-L-U, B-\sum_{j\in U} c_j y_j^{max}) < 0$. By (14) this is equivalent to

$$\frac{c_k}{\theta_k}\sum_{r\in K-U-L}\theta_r + \sum_{r\in U} y_r^{max} c_r - \sum_{r\in K-L-U} c_r > B - f,$$

and as B is increased, k will leave L when the inequality above no longer holds. The first k to leave L will be the one for which the expression of the left-hand side is the smallest. A similar condition is easily derived for when k which is currently not in U will enter U. This leads to the following expression:

$$\begin{array}{l}\bar{B} = \min\{\min_{k\in L}\{f + \frac{c_k}{\theta_k}\sum_{r\in K-U-L}\theta_r + \sum_{r\in U} y_r^{max}c_r - \sum_{r\in K-L-U} c_r\}, \\ \min_{k\in K-L-U}\{\frac{(y_k^{max}+1)c_k}{\theta_k}\sum_{r\in K-U-L}\theta_r + \sum_{r\in U} y_r^{max}c_r - \sum_{r\in K-L-U} c_r + f\}\}\end{array}. \tag{20}$$

Similarly, the lower endpoint $\underline{B}$ occurs when some $k \in U$ leaves U or some $k \in K-L-U$ enters L as B is reduced, resulting in the following condition.

$$\begin{array}{l}\underline{B} = \max\{\max_{k\in K-L-U}\{\frac{c_k}{\theta_k}\sum_{r\in K-U-L}\theta_r + \sum_{r\in U} y_r^{max}c_r - \sum_{r\in K-L-U} c_r + f\}, \\ \max_{k\in U}\{\frac{(y_k^{max}+1)c_k}{\theta_k}\sum_{r\in K-U-L}\theta_r + \sum_{r\in U} y_r^{max}c_r - \sum_{r\in K-L-U} c_r + f\}\}\end{array}. \tag{21}$$

To identify all breakpoints of $A(B)$ the following procedure can be followed. First, note that the relevant range of $B = [f, \sum_{k\in K} c_k y_k^{max}]$, since at the lower endpoint the optimal solution is $y_k = 0, k \in K$ and at the upper endpoint it is $y_k = y_k^{max}, k \in K$. We thus initialize $B_0 = f$ and the set of breakpoints $\mathbf{B} = \{f, \sum_{k\in K} c_k y_k^{max}\}$, choose some small $\epsilon > 0$, set $m = 0$ and proceed as follows:

Algorithm 2

1. Set $m = m+1, B = B_{m-1} + \epsilon$, and apply Algorithm 1 to find the optimal solution and the set L, U corresponding to the current B.

2. Use (21) to find the lower breakpoint $\underline{B}$. If $\underline{B} \neq B_{m-1}$, the step size was too large. Set $\epsilon = \epsilon/2$ and repeat Step 1. Else, proceed to Step 3.

3. Use (20) to find the next breakpoint B_m. If $B_m = \sum_{k\in K} c_k y_k^{max}$, stop: all budgetary breakpoints have been found. Else, add B_m to $\mathbf{B}$ and repeat Step 1.

The parametric analysis is summarized in the following result:

Theorem 2 *There exist a finite set of increasing breakpoints $B_1, \ldots, B_M$, with $B_1 = f$ and $B_M = \sum_{k\in K} c_k y_k^{max}$ such that*

(1) For $m = 1, \ldots, M-1$ and $B \in [B_m, B_{m+1}]$, the optimal value $A(B)$ of the objective

function of (SFDP) is given by (18). Moreover, the number of breakpoints $M \leq 2|K|$.

(2) The function $A(B)$ *is non-decreasing and continuous for* $B \in (0, B_M]$

(3) If $\sum_{k \in K} \theta_k \leq 1$ *(i.e., Assumption 2 holds), then* $A(B)$ *is concave for* $B \in [B_1, B_M]$.

Proof: Parts (1) and (2) follow directly from the discussion preceding the Theorem (continuity follows since the value of $A(B)$ as $B \to B_m$ is the same whether the breakpoint B_m is approached from above or below). The inequality $M \leq 2|K|$ follows by observing that each $k \in K$ can go through at most two transitions (once to leave L and once to enter U).

To prove part (3), first note that, as discussed earlier, $A(B)$ is concave on the interval between any two breakpoints, and thus is piece-wise concave for $B \in [B_1, B_M]$. It remains to prove that $A'(B_m)^+ \geq A'(B_m)^-$ for any breakpoint $B_m, m \in \{1, \ldots, M\}$, where $A'(B)^+, A'(B)^-$ are left and right derivatives, respectively.

Consider some breakpoint $B_m, 1 < m < M$ and let L_m, U_m, K_m be the sets L, U, K corresponding to $B \in [B_{m-1}, B_m]$. From the discussion above we know that there are two cases as the breakpoint B_m is crossed: (a) $L_{m+1} \subset L_m$ i.e., some variable that was pegged to its lower bound of 0 is unpegged, and (b) $U_m \subset U_{m+1}$ i.e., some variable that was not pegged is now pegged to its upper bound (both cases can occur simultaneously). Let $A_m(B)$ and $A_{m+1}(B)$ be the corresponding forms of the function $A(B)$ defined by (19) - recall that the sets L, U, K affects both the constants and the exponents in this definition. Note that

$$A(B) = \begin{cases} A_m(B) & \text{for } B \in [B_{m-1}, B_m] \\ A_{m+1}(B) & \text{for } B \in [B_m, B_{m+1}] \end{cases}$$

with $A_m(B_m) = A_{m+1}(B_m)$ by continuity of $A(B)$. Suppose case (a) holds. Then for any $B \in [B_{m-1}, B_m]$ we must have $A_{m+1}(B) \geq A_m(B)$ since there is one or more variable pegged to 0 under L_m that is unrestricted under L_{m+1}. Since both $A_m(B)$ and $A_{m+1}(B)$ are concave and increasing, and have the same value at B_m, It follows that $A'_m(B_m) \geq A'_{m+1}(B_m)$. Now, $A'(B_m)^+ = A'_m(B_m) \geq A'_{m+1}(B_m) = A'(B_m)^-$, establishing concavity of $A(B)$ at B_m.

Similarly, if case (b) holds then for $B \in [B_m, B_{m+1}]$ we must have $A_{m+1}(B) \leq A_m(B)$ since one or more variables that are pegged in U_{m+1} are unrestricted in U_m. Thus $A_m(B) \geq$

$A_{m+1}(B)$ must hold for all $B \in [B_m, B_{m+1}]$. As before, since both functions are concave and increasing, this implies that $A'_m(B_m) > A'_{m+1}(B)$. Thus $A'(B_m)^- = A'_{m+1}(B_m) \leq A'_m(B_m) = A'(B_m)^+$, which completes the proof.

$\square$

We illustrate the parametric analysis with the following example.

Example 2. Continuing with the setting of Example 1, we now wish to identify the set of budgetary breakpoints $\mathbf{B}$. We apply Algorithm 2 with $\epsilon = .01$. Since $f = 0$ and $\sum_k c_k y_k^{max} = .5 + 1 + 2 = 3.5$, the relevant search range is $B \in [0, 3.5]$ (for $B > 3.5$, all $y_k = y_k^{max} = 1$).

We start with $\mathbf{B} = \{0\}$, set $B_1 = \epsilon$ and apply Algorithm 1. This results in the optimal solution: $(y_1, y_2, y_3) = (.02, 0, 0)$ with $L = \{2, 3\}$ and $U = \emptyset$. Next we compute $\bar{B} = .05, \underline{B} = 0$ using (20), (21), respectively. Since $\underline{B} = 0$, i.e., the previous breakpoint, no breakpoints have been missed by taking too large a step. We thus add the upper breakpoint to the set $\mathbf{B} = \{0, 0.5\}$, set $B_2 = 0.5 + \epsilon$ and preform another iteration of Algorithm 2. The new optimal solution is $(y_1, y_2, y_3) = (1, 0.01, 0)$ with $L = \{3\}, U = \{1\}$ (note that node 1 entered U, while node 2 simultaneously left L). Here $\underline{B} = .05$, so no breakpoints have been missed and $\bar{B} = 1.5$, leading to $\mathbf{B} = \{0, .5, 1.5\}$. At the next iteration with $B_3 = 1.5 + \epsilon$ we obtain $(y_1, y_2, y_3) = (1, 1, 0.005)$, $L = \emptyset$, $U = \{1, 2\}$, with $\underline{B} = 1.5, \bar{B} = 3.5$. Since the upper breakpoint has reached the limit of the relevant range, the algorithm stops with the final set of breakpoints $\mathbf{B} = \{0, .5, 1.5, 3.5\}$. Note that for each budgetary range we have also computed the corresponding sets L, U, enabling us to evaluate $A(B)$ for any $B \geq 0$.

Remark 1 *It is important to note the role of Assumption 2. It is not necessary for the derivation of the optimal solution to the (SFDP) presented in Algorithm 1, or equations (18) and (19), which form the basis of the parametric analysis. Algorithm 2 also remains valid yielding the function $A(B)$ that is continuous, non-decreasing and differentiable except at the breakpoints when Assumption 2 does not hold. However, the concavity of $A(B)$, or even piece-wise concavity between the breakpoints, may be violated.*

4 Multi-Facility GFLDP

In this section we show how the (SFDP) solution methodology developed above can be applied to multi-facility (GFLDP) problem described earlier. Throughout this section we will assume that Assumption 2 holds, allowing us to treat the attractiveness $A_j(B_j)$ as a concave function of B_j in view of Theorem 2.

We start by formulating the problem as a non-linear integer program and then apply an extension of the TLA technique developed in [1] to develop an approximation with pre-specified level of relative error that can be obtained by solving a single linear integer program.

The primary decision variables are: $x_j, j \in P$, with the value of $x_j = 1$ indicating that a facility is opened at j and $x_j = 0$ otherwise. In addition, we use continuous decision variables $B_j \geq 0$, to represent the design improvement budget and $U_{ij} \geq 0, i \in N$ to represent the utility of facility at j to customers i. We also use parameter $g_{ij} = g(d_{ij})$ for $i \in N, j \in P$. Two version of the problem, depending on the type of the aggregator U_i used, are presented. We start with the $U_i = U_i^A$ defined by (2); recall that this aggregator is used to represent partial capture models. We call the corresponding formulation (GFLDP-A):

$$\max Z = \sum_{i \in N} w_i V_i(U_i) \tag{22}$$

Subject to

$$\sum_{j \in P} B_j \leq b \tag{23}$$

$$\sum_{j \in P} x_j \leq m \tag{24}$$

$$x_j f_j \leq B_j \leq x_j(f_j + \sum_k c_k y_k^{max}) \qquad j \in P \tag{25}$$

$$U_{ij} = A_j(B_j) g_{ij} \qquad i \in N, j \in P \tag{26}$$

$$U_i = \sum_{j \in P} U_{ij} + U_i^C \qquad i \in N \tag{27}$$

$$B_j \geq 0, U_{ij} \geq 0, x_j \in \{0, 1\} \qquad i \in N,\ j \in P \tag{28}$$

The objective function (22) is given by (4). Constraint (23) and (24) place upper bounds on the total budget and on the number of facilities, respectively. The next constraint 27) defines the utility U_i for customers at $i \in N$ (note that the values U_i^C are parameters that can be pre-computed). Constraint (25) ensures that $f_j + \sum_k c_k \geq B_j \geq f_j$ when $x_j = 1$ (i.e. B_j is large enough to cover the fixed construction costs, but does not exceed the maximum improvement budget). On the other hand, if $x_j = 0$ then this constraint forces $B_j = 0$. Constraint (26) defines the term U_{ij} of the utility vector for $j \in P, i \in N$ using the definition (1) and the expression (18) derived in the previous section. Note that $A_j(0) = 0$ and thus the utilities will automatically be forced to 0 if no facility is open at j.

Next we formulate (GFLDP) with the $U_i = U_i^M$ aggregator defined by (3) and used to represent full capture models. The corresponding formulation, referred to as (GFLDP-M), differs from the one above only with respect to constraint (27) which is now replaced with the following four constraints:

$$U_i \leq U_{ij} \qquad i \in N, j \in P \tag{29}$$

$$U_i - U_i^C R_i \geq 0 \qquad i \in N \tag{30}$$

$$U_i \leq R_i M \qquad i \in N \tag{31}$$

$$R_i \in \{0, 1\} \qquad i \in N \tag{32}$$

Here $R_i, i \in N$ is a binary decision variable and M is a sufficiently large constant. Observe that if for $i \in N$ we have $\max_{j \in P} U_{ij} \geq U_i^C$ then the solution $U_i = \max_{j \in P} U_{ij}$ and $R_i = 1$ is feasible. Moreover, since the objective is increasing in U_i, this solution will be optimal. If, on the other hand $\max_{j \in P} U_{ij} < U_i^C$, then the only feasible solution is $U_i = 0, R_i = 0$. Thus R_i can be interpreted as an indicator of whether customer i is captured by one of the new facilities. Selecting $M = \max_{j \in P} \alpha_j \prod_{k \in K} y_k^{max}$ will ensure that constraint (31) does not limit the value of U_i when $R_i = 1$.

Formulations (GFLDP-A) and (GFLDP-M) have non-linearities in both the objective and the constraints. However, all non-linearities involve concave functions. In the next section we show how concavity can be used to obtained a piece-wise linear approximations of the models above.

4.1 Piece-wise Linear Approximation

Our goal in this section is to develop a piece-wise linear "approximation scheme", that is a methodology for approximating non-linear terms in the formulations above with piece-wise linear functions where the error bound is specified by the user. This will, in term, allow us to develop linear Mixed Integer Programming formulations for our models that are accurate to within a pre-specified tolerance level.

We follow the approach based on Tangent Line Approximation (TLA) developed by Aboolian *et al.* in [1]. The main result we use is as follows:

Theorem 3 (from [1]) *Consider* $z(\mathbf{u}) = \sum_{i \in N} w_i v(u_i)$*, where* $v(u_i) \geq 0$ *is a concave, non-decreasing function of* $u_i \in [u^{min}, u^{max}]$ *for* $i \in N$ *and* $0 \leq u^{min} < u^{max}$ *are constants. Then*

1. *For a specified* $\epsilon > 0$ *the TLA algorithm computes a concave piece-wise linear function* $v^{\epsilon}(u), u \in [u^{min}, u^{max}]$ *such that*

$$v(u) \leq v^{\epsilon}(u) \leq (1+\epsilon)v(u) \text{ for all } u \in [u^{min}, u^{max}], \tag{33}$$

and the number of linear segments in v^{ϵ} *is minimal for all piece-wise linear functions satisfying the inequality above.*

2. *Inequality* $z(\mathbf{u}) \leq z^{\epsilon}(\mathbf{u}) := \sum_{i \in N} v^{L}(u_i) \leq (1+\epsilon)z(\mathbf{u})$ *holds for all* $\mathbf{u} \in [u^{min}, u^{max}]^n$*. Note that* $z^{\epsilon}(\mathbf{u})$ *is a piece-wise linear function.*

In words, the TLA procedure constructs a concave, piecewise linear upper-approximator $v^{\epsilon}(u)$ for each concave component that meets the specified maximum error ϵ, and allows us to replace the original $n-$dimensional non-linear function $z(\mathbf{u})$ with a piece-wise linear upper-approximator $z^{\epsilon}(\mathbf{u})$ that has the same maximum error. We will refer to $v^{\epsilon}(u)$ as the

"ϵ-level TLA-approximator" of $v(u)$. Since representing a piece-wise linear function in an MIP formulation requires a new decision variable for every linear segment, the fact that the TLA approximator has minimal number of segments ensures that it is the most efficient upper-approximator for the specified tolerance level ϵ.

Applying the previous result to the objective function of (GFLDP-A) and (GFLDP-M) formulations, allows us to obtain piece-wise linear objectives. However, non-linearities remain in the constraints since each U_i is itself a non-linear function of the budget variables $B_j, j \in S$. Observing that the terms U_i are expressible as linear functions of $A(B_j), j \in N$, and that, by Theorem 2 (assuming Assumption 2 holds), $A(B_j)$ is a concave function of B_j for each $j \in N$, a natural idea is to apply the TLA procedure iteratively: first to construct piece-wise linear approximators $A^{\epsilon}(B_j), j \in N$, yielding piece-wise linear $U_i^{\epsilon}(\mathbf{B})$, and then to each $V_i(U), i \in N$.

However, this "direct" approach may lead to violation of the specified error bound ϵ for the piece-wise linear approximator of $V_i(U)$ since it evaluates its argument at the "wrong" place (at U_i^{ϵ} instead of the true value U_i). The "correct" iterative TLA approach is developed in the following result.

Theorem 4 *For a specified error tolerance* $\epsilon > 0$ *let* $\phi = \sqrt{1+\epsilon} - 1$ *and constants* $0 \leq u^{min} < u^{max}$, *consider* $z(\mathbf{b}) = \sum_{i \in N} v_i\left(u_i(\mathbf{b})\right)$, *where for each* $i \in N$,

- $v_i(u) \geq 0$ *is a concave, non-decreasing function of* $u \in [u^{min}, u^{max}]$, *and*

- $u_i(\mathbf{b})$ *is a non-decreasing function of* $\mathbf{b} = (b_1, \dots, b_n), b_i \in [0, B]$ *with the range* $u_i(\mathbf{b}) \in [u^{min}, u^{max}]$.

Suppose piece-wise linear approximators v_i^{ϕ} *and* u_i^{ϕ} *are available such that for all* $i \in N$,

$$v_i(u) \leq v_i^{\phi}(u) \leq (1+\phi) v_i(u) \text{ for } u \in [u^{min}, u^{max}] \text{ ,and} \tag{34}$$

$$u_i(\mathbf{b}) \leq u_i^{\phi}(\mathbf{b}) \leq (1+\phi) u_i(\mathbf{b}) \text{ for all } \mathbf{b} \in [0, B]^n. \tag{35}$$

Then for all $i \in N$ and $\mathbf{b} \in [0, B]^n$,

$$v_i(u_i(\mathbf{b})) \le v_i^\phi(u_i^\phi(\mathbf{b})) \le (1+\epsilon)v_i(u_i(\mathbf{b})), \text{ and} \tag{36}$$

$$z(\mathbf{b}) \le z^\epsilon(\mathbf{b}) := \sum_{i \in N} v_i^\phi(u_i^\phi(\mathbf{b})) \le (1+\epsilon)z(\mathbf{b}). \tag{37}$$

Note that $z^\epsilon(\mathbf{b})$ is a piece-wise linear function of $\mathbf{b} \in [0, B]^n$.

Proof: For $i \in N$ and $\mathbf{b} \in [0, B]^n$, by (34) we have

$$v_i(u_i^\phi(\mathbf{b})) \le v_i^\phi(u_i^\phi(\mathbf{b})) \le (1+\phi)v_i(u_i^\phi(\mathbf{b})). \tag{38}$$

Now, by (35) and since $v_i(u)$ is non-decreasing,

$$v_i(u_i(\mathbf{b})) \le v_i^\phi(u_i^\phi(\mathbf{b})) \le v_i((1+\phi)u_i(\mathbf{b})).$$

This establishes the first inequality in (36). Also, since $v_i(u)$ is concave and non-decreasing, it follows that $v_i((1+\phi)u_i(\mathbf{b})) \le (1+\phi)v_i(u_i(\mathbf{b}))$. Thus, using (38),

$$v_i^\phi(u_i^\phi(\mathbf{b})) \le (1+\phi)v_i(u_i^\phi(\mathbf{b})) \le (1+\phi)^2 v_i(u_i(\mathbf{b})),$$

which establishes the second inequality in (36) since $(1+\phi)^2 = (1+\epsilon)$. The relationship (37) now follows directly, and the observation that $z^\epsilon(\mathbf{b})$ is piece-wise linear stems from the fact that each term is a composition of two piece-wise linear functions.

□

The previous result allows us to develop linearized versions of (GDFLP-A) and (GDFLP-M) formulations by (1) applying the TLA procedure to obtain estimator $V_i^\phi(U)$ for each term in the objective function, and (2) applying TLA procedure to each $A_j(B)$ to obtain estimator $A_j^\phi(B), j \in P$ which also leads to the piece-wise linear estimator for $U_i, i \in N$.

The approach is implemented as follows. For a specified maximum error $\epsilon > 0$ we let $\phi = \sqrt{1+\epsilon}-1$ and apply the TLA procedure to obtain $A_j^\phi(B_j)$ for $B_j \in [f_j, f_j + \sum_{k\in K} c_k y_k^{max}]$ and $j \in P$. Note that $A_j^\phi(B_j)$ consists of $r(j) \geq 1$ line segments, which are defined by breakpoints $b_j^1 := f_j < b_j^2 \ldots < b_j^{r(j)} < b_j^{r(j)+1} := f_j + \sum_{k\in K} c_k y_k^{max}$ and slopes $h_{A,j}^1 > h_{A,j}^2 > \ldots > h_{A,j}^{r(j)}$. When $B_j \in [b_j^r, b^(r+1)_j]$ for some $r \in 1, \ldots, r(j)$, the approximator $A^\phi(B)_j$ can be written as

$$A_j^\phi(B_j) = A_j(b_j^r) + (B_j - b_j^r)h_{A,j}^r, \tag{39}$$

where constants $b_j^r, h_{A,j}^r$, and $A_j(b_j^r)$ can all be computed during the pre-processing stage.

This form, however, cannot be used directly in our formulation for two reasons. First, we need an indicator variable to identify which subsegment the argument B_j belongs to. Second, recalling that $A_j(B_j)$ represents the attractiveness of facility j when budget B_j is allocated to that facility and that the attractiveness should be 0 when $B_j = 0$, we need to make sure that our approximator also evaluates to 0 in this case. To that end, we let $\ell_{A,j}^r = b^{(r+1)_j} - b_j^r$ be the length of segment $r \in 1, \ldots, r(j)$ and rewrite (39) as follows:

$$A_j^\phi(B_j) = A_j(b_j^r)W_{A,j}^r + \ell_{A,j}^r h_{A,j}^r Q_{A,j}^r, \tag{40}$$

where $W_{A,j}^r, Q_{A,j}^r$ are the new decision variables defined by

$$W_{A,j}^r = \begin{cases} 1 & \text{if } B_j \in [b_j^r, b_j^{(r+1)}] \\ 0 & \text{otherwise,} \end{cases} \qquad Q_{A,j}^r = \begin{cases} (B_j - b_j^r)/\ell_{A,j}^r & \text{if } W_{A,j}^r = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{41}$$

In words, $W_{A,j}^r$ indicates which segment $r \in 1, \ldots, r(j)$ the argument B_j belongs to and $Q_{A,j}^r$ measures the proportion of that segment that is covered. The advantage of (40) is that it evaluates to 0 when $W_{A,j}^r = Q_{A,j}^r = 0$. This allows us to replace constraints (26, 27) in formulation (GFLDP-A) with the the following constraints:

$$B_j = \sum_{r=1}^{r(j)} [b_j^r W_{A,j}^r + \ell_{A,j}^r Q_{A,j}^r] \qquad j \in P \qquad (42)$$

$$\sum_{r=1}^{r(j)} W_{A,j}^r \leq 1, \qquad j \in P \qquad (43)$$

$$W_{A,j}^r \geq Q_{A,j}^r, \qquad j \in P, r \in \{1, \ldots, r(j)\} \qquad (44)$$

$$U_{ij}^\phi = g_{ij} \sum_{r=1}^{r(j)} [A_j(b_j^r) W_{A,j}^r + \ell_{A,j}^r h_{A,j}^r Q_{A,j}^r] \qquad i \in N, j \in P \qquad (45)$$

$$U_i = \sum_{j \in P} U_{ij}^\phi + U_i^C \qquad i \in N \qquad (46)$$

$$W_{A,j}^r \in \{0, 1\},\ Q_{A,j}^r \geq 0, U_{ij}^\phi \geq 0 \qquad j \in P, r \in \{1, \ldots, r(j)\} \qquad (47)$$

Constraint (42) represents the budget B_j allocated to facility j in terms of the new decisions variables $W_{A,j}^r, Q_{A,j}^r$. This, together with the next two constraints enforces the definitions (41). Indeed, suppose for some $j \in P$ the constraint (25) allows B_j to take on a positive value (i.e., $x_j = 1$ signifying an open facility at j). Then, from (25) and the definitions of breakpoints above, we must have $B_j \in [b_j^r, b_j^{r+1}]$ for some $r \in \{1, \ldots, r(j)\}$. Defining $W_{A,j}^r, Q_{A,j}^r$ according to (41) clearly results in values that make constraints (42-44) feasible. To see that these values are unique, note that (43, 44) allow $W_{A,j}^r, Q_{A,j}^r$ to be non-zero for at most one value of r and there is clearly only one way to represent $B_j > 0$ in (42). On the other hand, if $x_j = 0$ (i.e., no facility is open at j), then $B_j = 0$ by (25), which forces $W_{A,j}^r = Q_{A,j}^r = 0$ for all $r = 1, \ldots, r(j)$. Constraint (45) now defines the approximate value U_{ij}^ϕ of the utility term; its correctness following directly from (26) and (40). Note that the new decision variables U_{ij}^ϕ replace the variables U_{ij} in the original formulation. The approximate customer-level utility U_i is now defined by (46).

We have now linearized all constraints of (GFLDP-A). To linearize the objective function we apply the same process to each term $V_i(U)$ for $i \in N$. First observe that the maximum utility $\bar{U}_i = U_i^C + \sum_{j \in P} A_j(b)(1 + d_{ij})^{-\beta}$ when the average aggregator U_i^A is used, and $\bar{U}_i = U_i^C + \max_{j \in P} A_j(b)(1 + d_{ij})^{-\beta}$ when then U_i^M aggregator is used. The minimum value for U_i is U_i^C in both cases. Thus, the relevant range for the argument of $V_i(U)$ is $U \in [U_i^C, \bar{U}_i]$. Applying the TLA process with tolerance level ϕ we obtain $\rho(i)$ breakpoints $u_i^1 := U_i^C < u_i^2 < \ldots < u_i^{\rho(i)} < u_i^{\rho(i)+1} := \bar{U}_i$ with the corresponding set of non-increasing slopes $h_{V,i}^r,\ r = 1, \ldots, \rho(i)$. Letting $\ell_{V,i}^r = u_i^{(r+1)} - u_i^r$ be the length of segment r and defining new decision variables $W_{V,i}^r, Q_{V,i}^r$ by

$$W_{V,i}^r = \begin{cases} 1 & \text{if } U_i \in [u_i^r, u_i^{(r+1)}] \\ 0 & \text{otherwise,} \end{cases} \qquad Q_{V,i}^r = \begin{cases} (U_i - u_i^r)/\ell_{V,i}^r & \text{if } W_{V,i}^r = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{48}$$

we can write

$$\begin{aligned} U_i &= \sum_{r=1}^{\rho(i)} [u_i^r W_{V,i}^r + \ell_{V,i}^r Q_{V,i}^r], \\ V_i^\phi(U_i) &= V_i(u_i^r) W_{V,i}^r + \ell_{V,i}^r h_{V,i}^r Q_{V,i}^r, \end{aligned}$$

with constants $u_i^r, h_{V,i}^r, \ell_{V,i}^r, V_i(u_i^r)$ computable during the pre-processing stage for each $i \in N$ and $r \in \{1, \ldots, \rho(i)\}$.

We now add constraints

$$U_i = \sum_{r=1}^{\rho(i)} [u_i^r W_{V,i}^r + \ell_{V,i}^r Q_{V,i}^r] \qquad i \in N \tag{49}$$

$$V_i^\phi = \sum_{r=1}^{\rho(i)} [V_i(u_i^r) W_{V,i}^r + \ell_{V,i}^r h_{V,i}^r Q_{V,i}^r] \qquad i \in N \tag{50}$$

$$\sum_{r=1}^{\rho(i)} W_{V,i}^r = 1 \qquad i \in N \tag{51}$$

$$W_{V,i}^r \geq Q_{V,i}^r \qquad i \in N, r \in \{1, \ldots, \rho(i)\} \tag{52}$$

$$W_{V,i}^r \in \{0,1\},\ Q_{V,i}^r \geq 0, \qquad i \in N, r \in \{1, \ldots, \rho(i)\} \tag{53}$$

to formulation (GFLDP-A) and replace the objective function (22) with the linear objective

$$Z^{\epsilon} = \sum_{i \in N} w_i V_i^{\phi}, \tag{54}$$

obtaining a linear formulation. To summarize, previous discussion together with Theorem 4 leads to the following lineat IP formulation for (GFLDP-A):

Corollary 1 *Consider proportional allocation (GFLDP) with average utility aggregator* U_i^A *given by (2). For a given error tolerance* $\epsilon \in (0, 1)$ *consider the Integer Program (IP) with objective* Z^{ϵ} *given by (54) and constraints (23),(24),(28),(42-47), (49-53). The optimal value of this IP has the maximum relative error of* ϵ *with respect to the optimal value of (GFLDP-A).*

To obtain a similar result for the full capture version of (GFLDP) we only need to replace constraint (46) with

$$U_i \leq U_{ij}^{\phi}, \ i \in N \tag{55}$$

and use the formulation (GFLDP-M), leading to the following result.

Corollary 2 *Consider full capture (GFLDP) with maximum utility aggregator* U_i^M *given by (3). For a given error tolerance* $\epsilon \in (0, 1)$ *consider the Integer Program (IP) with objective* Z^{ϵ} *given by (54) and constraints (23),(24),(28),(30-32), (42-45), (55),(47), and (49-53). The optimal value of this IP has the maximum relative error of* ϵ *with respect to the optimal value of (GFLDP-M).*

A few remarks are in order with respect to these results. First, in view of the discrete nature of the solution space, it is trivial to observe that by setting ϵ small enough we can guarantee that the solution obtained by our approximating linear IPs in the previous two results are (exactly) optimal with respect to the corresponding models, though, of course, the smaller the ϵ, the more linear segments are required to achieve the desired degree of accuracy in the TLA approximators for $A_j(B)$ and $V_i(U)$, which increases the dimensionality of the resulting IPs (since new segments require new decision variable $W^r_{A,j}, Q^r_{A,j}, W^r_{V,i}, Q^r_{V,i}$). On the other hand, as previously observed in [1], the number of segments in the TLA approximation tends to grow slowly with ϵ, making high level of accuracy (or possibly optimal solutions) achievable. We also note that the error tolerances represent the em worst-case maximum errors; the errors observed in experiments are typically much smaller. in fact, we find that $\epsilon = 5\%$ typically leads to optimal solutions for most instances where the exact optimal solution is known.

Second, while the same error tolerance $\phi = \sqrt{1+\epsilon} - 1$ was used to derive TLA approximators of both $A_j(B)$ and $V_i(U)$, this is not necessary - any values ϕ_1, ϕ_2 such that $(1+\phi_1)(1+\phi_2) = \epsilon$ can be used in Theorem 4. Thus, if it turns out that some non-linear components are substantially more difficult to approximate than others, the tolerance levels could be increased for "tougher" components and decreased for "easier" ones.

Example: We illustrate the results above with their application to the classic Uncapacitated Facility Location Problem (UFLP), which is identical to MAXCAP model of [15] when no competition is present (i.e., $U^C_i = 0$ for all $i \in N$). Since all functions are linear in this case, we expect (GDFLP) to recover the standard UFLP formulation. As the following discussion shows, this is indeed the case.

As discussed in Section 2.2 above, in this case $V_i(U) = U, g_{ij} = g(d_{ij}) = d^{-1}_{ij} = r_{ij}, f_j = 1, K = \emptyset, A_j(B) = 1$for $B \geq 1$ for all $j \in P, i \in N$ and that $b = m$. Moreover, in the simple version of the model there are no competitors either, implying $U^C_i = 0$ for all $i \in N$.

We start with the (GFLDP-M) formulation. Observe that constraint (25) implies that $B_j = x_j$ and (23, 24) are equivalent. Thus,we can drop variables x_j and constraint (24), replacing them with a simple $B_j \in \{0, 1\}$ condition (i.e., variable B_j becomes an indicator of whether a facility is located at $j \in P$. Moreover, since $U^C_i = 0$ for all $i \in N$, constraints (30-32), together with variables $R_i, i \in N$ can be dropped as well; they are replaced with a simple non-negativity condition $U_i \geq 0, i \in N$.

Next, we observe that since $A_j(B) = 1$ for $B \geq 1$ (for any j), the TLA procedure trivially returns $r(j) = 1, b^1_j = 1, b^2_j = m, A_j(b^1_j) = 1, h^1_{A,j} = 0, \ell^1_{A,j} = (m-1)$ (i.e., for any $\phi > 0$, the piece-wise linear approximator $A_j^{\phi}(B)$ consists of one sub-segment of length $m-1$ and has intercept of 1 and slope of 0). Constraints (42-45), (55),(47) therefore become

$$
\begin{aligned}
B_j = W_{A,j}^1 + (m-1)Q_{A,j}^1] & \qquad j \in P \\
W_{A,j}^1 \leq 1, & \qquad j \in P \\
W_{A,j}^1 \geq Q_{A,j}^1, & \qquad j \in P, \\
U_{ij}^{\phi} = r_{ij} W_{A,j}^1 & \qquad i \in N, j \in P \\
U_i \leq U_{ij}^{\phi}, & \qquad i \in N W_{A,j}^1 \in \{0,1\},\ Q_{A,j}^1 \geq 0, U_{ij}^{\phi} \geq 0 \qquad j \in P.
\end{aligned}
$$

It is clear that variables $Q_{A,j}^1$ play no role in determining U_{ij}^{ϕ} and can be removed. Moreover, since (as observed previously) B_j is binary, as is $W_{A,j}^1$, and the first constraint above implies $B_j = W_{A,j}^1$. Therefore variables $W_{A,j}^1$ can be removed as well and all of the above constraints are replaced with

$$
\begin{aligned}
U_{ij}^{\phi} = r_{ij} B_j, & \qquad i \in N, j \in P \\
U_i \leq U_{ij}^{\phi}, & \qquad i \in N, U_{ij}^{\phi} \geq 0, U_i \geq 0 \qquad i \in N, j \in P.
\end{aligned}
$$

Finally, we examine the constraints (49-53) and the objective (54). Since the demand function $V_i(U) = U$ is already linear and $U \in [0, \bar{U}]$, where the upper bound $\bar{U}$ can be defined as some sufficiently large number (e.g., $\bar{U} = \max_{i,j} r_{ij}$, we observe that for any $\phi > 0$ and $i \in N$ we have $\rho(i) = 1, u_i^1 = 0, u_i^2 = \bar{U}, h_{V,i}^1 = 1, V(u_i^1) = 0, \ell_{V,I}^1 = \bar{U}$. Thus we obtain new constraints

$$
\begin{aligned}
U_i = 0 \cdot W_{V,i}^1 + \bar{U} Q_{V,i}^1 & \qquad i \in N \\
W_{V,i}^1 = 1, & \qquad i \in N \\
W_{V,i}^1 \geq Q_{V,i}^1, & \qquad i \in N, \\
V_i^{\phi} = 0 \cdot W_{V,i}^1 + 1 \cdot \bar{U} Q_{V,i}^1 & \qquad i \in N \\
W_{V,i}^1 \in \{0,1\},\ Q_{V,i}^1 \geq 0, V_i^{\phi} \geq 0 & \qquad i \in N,
\end{aligned}
$$

which simplify to $V_i^\phi = U_i, i \in N$, with variables $W_{V,i}^1, Q_{V,i}^1$ dropping out entirely. The linearized objective now takes the form $Z^\epsilon = \sum_{i \in N} w_i V_i^\phi = \sum_{i \in N} w_i U_i$. Putting it all together we obtain

$$\begin{aligned} \max \sum_{i \in N} w_i U_i, & \\ \text{Subject to} & \\ \sum_{j \in P} B_j \leq m & \\ U_i \leq B_j r_{ij} & \qquad\qquad i \in N, j \in P \\ B_j \in \{0,1\}, U_i \geq 0, & \qquad\qquad i \in N, j \in P, \end{aligned}$$

which is equivalent to the standard UFLP formulation. Of course, our approach allows us to easily generalize this model by adding in non-identical facility costs, design decisions, and non-linear demand.

5 Computational Experiments

In this section we conduct a set of computational experiments to solve the (GFLDP) via the techniques developed in the previous sections. For our experiments e used a subset of networks from the well-known set of p-median test problems by Beasley [3] with number of nodes N ranging from 100 to 900; these instances provided the network structure, as well as node weights. We also used the following functional forms:

- Demand function $V(U) = 1 - \exp(-\lambda U)$. Note that $\lambda > 0$ can be interpreted as (in)elasticity of demand - for higher values of λ the demand is less elastic
- Distance function $g(d) = (1 + d)^{-\beta}$. Here $\beta > 0$ represents customer sensitivity to distance (higher values correspond to higher sensitivity)
- Utility aggregator $U_i^A = \sum_{j \in P} U_{ij}$

Since a common (and practical) approach to location and design problems is to solve them separately, i.e., first determine the locations and number of facilities and then determine the design for each location, we wanted to see to what extend this approach underperforms the solutions of GFDLP model that jointly optimizes locations and design. To that end we developed the following heuristic "HS" that, for a given number of facilities first finds "good" locations via the $m-$median model, and then the optimal design for each of these locations using the single facility design problem (SFDP):

Heuristic HS

Step 0 Let $c^{min} = \min_{j\in P} f_j$ and $c^{max} = \max_{j\in P}(f_j + \sum_{k\in K} c_k y_k^{max})$ be the minimum and maximum facility costs, respectively. Then $m^{min} = \lfloor b/c^{max} \rfloor$, $m^{max} = \lfloor b/c^{min} \rfloor$ are the minimum and maximum number of facilities that can be located for a given budget b.

Step 1 For m between m^{min} and m^{max} do

Step 1.1 Solve $m - median$ model to find locations for m "basic" (i.e., all design improvement characteristics set to 0) facilities.

Step 1.2 Allocate budget f_j to each facility j. In case b is not sufficient to open minimal facilities at all m locations, open facilities in order of increasing $m-$median objective values. Any leftover budget is allocated to facilities in inverse proportionality to the $m-$median values. Solve (SFDP) model for each location with non-zeor additional budget to determine the optimal facility design for that location.

Step 1.3 Compute the (GFLDP) objective corresponding to this solution and increment m.

Step 2 Output the solution with the highest objective value.

Results of computational experiments for single and multiple facility facility problems are presented in the next two sections.

5.1 Solving Single-Facility Problem - Computational Experiments

Computational results for SFDP presented below are based on Algorithms 1 and Algorithm 2 of Section 3. Since both GFLDP and the (HS) heuristic are limited to 1 facility, the difference in solutions is due to the fact that the former first computes optimal design for each node and then selects the best node-design combination, while the latter selects $1-$median as the facility location and then optimizes the design for this location. No pre-existing competitive facilities were assumed and the following parameter values were used:

- The number of design characteristics $|K|$ was set to $3, 10, 100$. For each $k \in K$, the sensitivity parameter θ_k in the attractiveness formula (5) was randomly selected from Uniform$[0, 1]$. We fixed $\alpha_j = 1, j \in N$ and $y_k^{max} = 1, k \in K$.
- The costs $c_k, k \in K$ were randomly drawn from Uniform$[1, 10]$ for each k. The budget was set to $b = 7.5|K|$, and the fixed costs were set to $f_j = (M^*/M_j)2.5|K|$ where M^* and M_j are the values of the optimal $1-$median solution and the $1-$median solution at $j \in N$, respectively. This ensured that there was some budget for facility design at each location, but that more attractive locations (with better $1-$median values) had higher fixed costs.
- The distance sensitivities β were set to $1.5, 2, 2.5$ and demand sensitivities λ to $.5, 1, 1.5$.

Altogether, 271 instances were generated. The basic results can be found on Table 5.1. Columns 3-5 contain the average run time (sec), the average number of breakpoints of the $A(B)$ function in Algorithm 2, and the ratio of objective values of (HS) and (SFDP) solutions.

First, observe that SFDP solution is extremely fast (for 100-node networks, the runtimes were essentially instantaneous). Second, the number of breakpoints of the $A(B)$ function appears to scale approximately linearly with the number of design characteristics. Finally, the average gap (HS) and (SFDP) solutions is about 12%. Further examination of results shows that this gap is strongly affected by two parameters: distance sensitivity β and demand in-elasticity λ; when β is high and λ is low the gaps are quite high larger - this is displayed

# Nodes	Design Characteristics |K|	Values Average of time to find optimal (sec)	Average of # of break points	Average of Zm/Zsfd	Average of Zm/Zsfd (high distance sensitivity and demand elasticity)
100	**3**	-	4.4	96%	84%
	10	-	19.0	91%	76%
	100	-	197.8	91%	78%
500	**3**	0.0166	4.4	84%	50%
	10	0.0192	19.0	84%	49%
	100	0.0277	197.8	84%	50%
900	**3**	0.0625	4.4	88%	56%
	10	0.0655	19.0	85%	47%
	100	0.0807	197.8	87%	56%
Grand Total		**0.0302**	**73.7**	**88%**	**61%**

Table 5.1: Computational results for SFDP model. Last column corresponds to values $\beta = 2.5, \lambda = .5$

in the last column of Table 5.1. This observation makes sense: when customers are not either sensitive to travel distance, or the demand is not elastic, the location of the facility will not have a strong effect on the intercepted demand, and the difference between HS and SFDP solutions will be small. However, when both, strong distance sensitivity and demand elasticity are present, separating location and design decisions may result in optimality gaps of over 50%.

5.2 Solving Multi-facility (GDFLP) - Computational Experiments

The basic setup of for the multi-facility experiments was similar to the one used in the previous section, with the following changes:

- P-median Beasley set instances 1 and 4 ($n = 100$) and 7 and 10 (n=200) were used.=
- The maximum number of facilities m was set to $5, 7, 10$ for $n = 100$ and $10, 15, 20$ for $n = 200$. The set of potential locations was the node set N
- The number of competitive facilities $|N^C|$ was set to $0, 1, 3, 5$. For the case with non-zero competitive facilities, the competitor was treated as a leader, i.e., we first solved

GFDLP (assuming no competition and same set-up with respect to design characteristics, costs, and budgetary constraints as for own facilities) to find optimal locations and design for competitive facilities, then fixed these facilities and located up to m own facilities.

- The number of design characteristics was set to $|K| = 3$ with $\theta_1 = .25, \theta_2 = .5, \theta_3 = .75$ and $y_k^{max} = 1$ for $k = 1, 2, 3$. This allows us to track the degree of "improvement" of each located facility. The base attractiveness level was set to $\alpha_j = 1$ for all j.

- The design improvement costs were set as follows: $c_k = \Gamma \bar{f} 2^{(\theta_k - 1)}$, where $\bar{f} = (1/n) \sum_{j \in N} f_j$ is the average fixed cost and Γ is the "design cost parameter" with values $\{0.1, ..., 0.5\}$. This cost structure was selected to ensure that (a) the characteristics with higher θ_k values (i.e., having more impact on attractiveness) are also more expensive to improve, and (b) design costs are scaled relative to to fixed costs. The parameter Γ allows us to make overall cost of design improvements more or less expensive relative to fixed location costs.

- Budget b was calibrated to be approximately equal to the cost of opening m "basic" facilities (i.e., facilities with all design improvement variables Y_k set to 0). This gave the model options to either locate up to m basic facilities, or a lesser number of facilities with improved design.

- All GFLDP instances were solved using the iterated TLA approach described in Section 4 with maximum error tolerance set to $\epsilon = 5\%$. All MIPs were solved using CPLEX solver with the time limit set to one hour. If convergence was not achieved within this time, the best found solution was reported.

- HS heuristic solution was also obtained for every problem instance.

Altogether, 4320 problem instances were solved. We report a summary of key results below; full results are available from authors upon request. Most observations reported below were confirmed with statistical analysis of results.

We start by analyzing runtimes (in seconds) for both GFLDP and HS solutions on Tables 5.2(a)-(d). First, we observe that the methodology developed above for GFLDP leads to

Nodes	Facilities	Average of GFLDP Run Time	Average of HS Run Time
100	5	10.68	9.27
	7	7.90	15.52
	10	38.63	34.38
100 Total		19.07	19.72
200	10	79.95	124.28
	15	73.55	239.48
	20	99.83	349.68
200 Total		84.44	237.81

Runtime table (a)

N	Comp. facilities	Average of GFLDP Run Time	Average of HS Run Time
100	0	21.54	20.69
	1	21.32	19.07
	3	16.59	20.87
	5	16.82	18.26
100 Total		19.07	19.72
200	0	95.69	306.64
	1	92.86	255.60
	3	78.90	217.60
	5	70.31	171.41
200 Total		84.44	237.81

Runtime table (b)

N	Design cost coeff. (Gamma)	Average of GFLDP Run Time	Average of HS Run Time
100	0.1	78.07	17.75
	0.2	6.40	20.57
	0.3	5.22	18.32
	0.4	3.20	23.70
	0.5	2.45	18.28
100 Total		19.07	19.72
200	0.1	190.45	229.59
	0.2	187.70	219.10
	0.3	30.30	235.47
	0.4	8.66	267.28
	0.5	5.10	237.62
200 Total		84.44	237.81

Runtime table (c)

N	CFLDP Status	Design cost coeff. (Gamma)	% of Cases	Count of Cases	Average of GFLDP Run Time	Average of HS Run Time
100	Current	0.1	0.25%	11	606.67	45.97
	Current Total		0.25%	11	606.67	45.97
	Optimal	0.1	9.75%	421	64.26	17.01
		0.2	10.00%	432	6.40	20.57
		0.3	10.00%	432	5.22	18.32
		0.4	10.00%	432	3.20	23.70
		0.5	10.00%	432	2.45	18.28
	Optimal Total		49.75%	2,149	16.06	19.59
100 Total			50.00%	2,160	19.07	19.72
200	Current	0.1	1.83%	79	611.03	213.23
		0.2	1.97%	85	614.64	12.49
	Current Total		3.80%	164	612.90	109.18
	Optimal	0.1	8.17%	353	96.33	233.25
		0.2	8.03%	347	83.12	269.72
		0.3	10.00%	432	30.30	235.47
		0.4	10.00%	432	8.66	267.28
		0.5	10.00%	432	5.10	237.62
	Optimal Total		46.20%	1,996	41.02	248.38
200 Total			50.00%	2,160	84.44	237.81

Runtime table (d)

Table 5.2: Analysis of runtimes (in sec) for GFLDP and HS.

runtimes that are, on average, significantly *smaller* than for the HS heuristic: while the difference is minor for $n = 100$ case, it increases to nearly four times for $n = 200$ networks. This effect was also verified on larger networks (results available on request). Table 5.2(a) shows that runtimes increase with the network size n and maximum number of facilities m for both GFLDP and HS. However, the second parameter has much stronger impact on HS runtimes vs. GFLDP. Table 5.2(b) shows that runtimes tend to decrease as the number of competitive facilities grows (again, the effect is stronger for HS). Table 5.2(c) illustrates one of the strongest factors affecting runtimes of GFLPD: the design cost parameter Γ (this effect was confirmed with statistical analysis). This is because higher values of Γ limit facility design options under the available budget, thus making the problem easier. On the other hand, this parameter has no impact on HS runtimes. As a result, instances with the lowest value of Γ were the only ones where average runtimes for HS were smaller than for GFLDP. Table 5.2(d) provides results for cases that could not be solved to optimality within the time limit - labeled as "Current" in the "GFLDP Status" column; there were 11 out of 2160 such instances for $n = 100$ case, and 164 out of 2160 instances for $n = 200$ case. It can be seen that all of these cases had low values of Γ; for $n = 200$ nearly 20% of instances with $\Gamma = 0.1, 0.2$ could not be solved to optimality - further confirming that relative cost of facility design is the key determinant of the computational difficulty of GDFLP instance. We also conclude that problems with up to 200 nodes and potential facility locations and up to 20 facilities can typically be solved quite quickly using the iterated TLA approach for the TLA.

N	GFLDP Status	Design cost coeff. (Gamma)	Average of HS Gap
100	Current	0.1	33%
	Current Total		**33%**
	Optimal	0.1	31%
		0.2	23%
		0.3	12%
		0.4	5%
		0.5	1%
	Optimal Total		**14%**
100 Total			**14%**
200	Current	0.1	31%
		0.2	16%
	Current Total		**23%**
	Optimal	0.1	26%
		0.2	25%
		0.3	16%
		0.4	8%
		0.5	2%
	Optimal Total		**15%**
200 Total			**15%**

Table 5.3: Analysis of HS Gap - the relative difference between best-found HS and GFLDP solutions.

We next turn our attention to solution quality, focusing on the relative difference between the best solutions obtained by GFLDP and HS for each problem instance; we call this difference the "HS Gap"; the results are displayed on Table 5.3. The average HS Gap is about 15% - pointing to significant advantages of jointly optimizing design and location decisions (in GFLDP) vs making this decisions separately (in HS). According to statistical analysis, the main predictor of HS Gap is the design cost parameter Γ (the other important predictors are the maximum number of facilities m and distance sensitivity β). The effect of Γ can be seen on Table 5.3 - the HS Gap approaches 30% for lower values of Γ, i.e., when the available budget allows for more variation in facility design. It is interesting to note that some of the largest HS Gaps were observed for cases that were not solved to optimality within the time limit - indicating that even when GFLDP solution process is interrupted due to time constraints, the best-found solution is better than the HS alternative.

Finally, Tables 5.4(a),(b) explore the reasons behind the difference between GFLDP solutions. Two measures are displayed: the number of facilities opened by each algorithm and the average percentage of the available budget that was spend on facility locations (measured as the ratio of the sum of fixed location costs for all open facilities to the available budget);

N	m	Average of GFLDP # Facilities Located	Average of HS # Facilities Opened	Average of GFLDP % Budget Spent on Location	Average of HS % Budget Spent On Location
100	5	4.25	6.80	71%	90%
	7	5.72	9.49	60%	93%
	10	6.98	13.19	54%	95%
100 Total		**5.65**	**9.83**	**62%**	**93%**
200	10	8.19	11.49	62%	91%
	15	9.33	16.62	48%	95%
	20	10.40	21.72	41%	97%
200 Total		**9.31**	**16.61**	**50%**	**94%**

Number of facilities analysis (a)

N	Design cost coeff. (Gamma)	Average of GFLDP # Facilities Located	Average of HS # Facilities Opened	Average of GFLDP % Budget Spent on Location	Average of HS % Budget Spent On Location
100	0.1	4.36	9.13	58%	91%
	0.2	4.00	9.91	50%	93%
	0.3	4.83	10.09	50%	93%
	0.4	6.90	9.96	60%	93%
	0.5	9.83	10.04	90%	94%
100 Total		**5.98**	**9.83**	**62%**	**93%**
200	0.1	8.88	15.18	58%	92%
	0.2	7.86	16.43	46%	94%
	0.3	7.80	17.11	40%	95%
	0.4	8.20	17.06	38%	95%
	0.5	13.81	17.25	70%	95%
200 Total		**9.31**	**16.61**	**50%**	**94%**

Number of facilities analysis (b)

Table 5.4: Analysis of the number of facilities in the optimal solution of GFLDP and HS algorithms.

note that the remaining portion of budget must have been spent on design improvements. Here the differences between GFLDP and HS solutions are quire dramatic: Table 5.4(a) shows that GFLDP locates about 50% fewer facilities than HS (note that m is an upper bound on the number of facilities for GFLDP but not for HS). The last two columns of the table show that while GFLDP splits the available budget almost evenly between fixed location and design improvement costs, HS tends to spend nearly all available budget on the former, and only 6% on the latter. Table 5.4(b) shows that deisgn cost parameter Γ is the key determinant of the percentage of budget spent on location vs improvement by GFLDP, while it has almost no impact on HS solution - thus even when opportunities for cost efficient design improvements exist, HS algorithm cannot take advantage of them. To summarize, joint optimization of location and design leads to a smaller number of "improved" facilities, while separate optimization leads to a large number of "basic" facilities.

6 Concluding Remarks and Directions for Future Research

In this paper we have developed a general GFLPDP model that allows us to represent a rich variety of location and design models, including models with elastic demand, pre-existing competitive facilities, all-or-nothing or proportional customer allocation mechanisms, and a different types of design dimensions. Many previously described location models, including such classical models as UFLP, MaxCap, Competitive Interaction - appear as special cases within our framework.

We also develop an efficient solution methodology for GFDLP which allows us to solve medium-size problems (with several hundred potential facility locations and up to 20 new facilities to be located) within reasonable times.

One of the key advantages of GDFLP is that it allows how joint optimization of location and design decisions - answering the question of "what kind of facilities to locate" in addition to "where" and "how many" that are traditionally addressed by location models. Our computational results show that this joint optimization may be quite important in practice - separating design and location decisions, even in a reasonably sophisticated way as represented by our HS heuristic, leads to significant optimality gaps.

We close with comments on several extensions of the model. Perhaps the most obvious is changing the objective to profit (net revenue) optimization instead of maximizing revenue for a given budget (which is assumed in the current version). In principle, making budget b a decision variable which is subtracted from the current objective leads to no technical difficulties at all - all of the methodological developments above extend to this profit-optimizing version. The major exception is the relative error guarantee in Theorem 4. The reason is quite clear: the iterated TLA approach bounds the relative error of the revenue term; however if the profit margins are slim, the optimal profit may be a quite small relative to revenue, and thus the relative error for the piece-wise linear approximation may be quite large. It would be nice to develop an approach that extends the guarantees on the overall relative error to the profit-maximizing version of GFLDP.

Another obvious (and difficult) extension is to allow for more dynamic competition. As is the case for much of discrete-network competitive location models, GFLDP essentially solved the follower's problem in a leader-follower game. The leader's problem is more difficult, but of obvious importance.

Finally, the extensions to more general cost structure and attractiveness function structure for the design problem should be considered.

References

[1] Aboolian, R., Berman, O., and D. Krass, “Competitive Facility Location Model with Concave Demand,” *European Journal of Operations Research*), 181(2), 2007, 598-619.

[2] Aboolian, R., Berman, O., and D. Krass, “Competitive Facility Location and Design Problem,” *European Journal of Operations Research*, 182(1), 2007, 40-62.

[3] Beasley, J.E. “OR-library: distributing test problems by electronic mail,” *J Oper Res Soc*, 41, 1990, 1069-1072.

[4] Berman, O., Drezner, T., Drezner, Z., and D. Krass,“Modeling Competitive Facility Location Problems: New Approaches and Results”, Invited chapter in *Tutorials Operations Research*, M. Oskoorouchi, ed., INFORMS, 156-181, 2009.

[5] Berman, O. and D. Krass, “Locating Multiple CompetitiveFacilities: Spatial Interaction Models with Variable Expenditures,” *Annals of Operations Research*, Vol. 111, 2002, pp. 197-225.

[6] Berman, O. and D. Krass, “Flow Intercepting Spatial Interaction Model: A New Approach to Optimal Location of Competitive Facilities,” *Location Science*, Vol. 6, 1998, pp. 41-65.

[7] Birtran, G., and Hax, A.C., Disaggregation and Resource Allocation Using Convex Knapsack Problems With Bounded Variables, *Management Science*, Vol. 27, 1981, pp 431-441.

[8] Bretthauer, K.M., and Shetty, B., The Non-linear Knapsack Problem - Algorithms and Applications, *Europ. Journal of Oper. Res.*, Vol. 138, 2002, pp 459-472.

[9] Bretthauer, K.M., and Shetty, B., The Non-linear Resource Allocation Problem, *Operations Research*, Vol. 43, 1995, pp 670-683.

[10] Ghosh, A., S.L. McLafferty and C.S. Craig, “Multifacility Retail Networks,” *Facility Location*, Z. Drezner, ed., Springer-Verlag, 1995, pp 301-330.

[11] Hakimi, S.L., " Locations with spatial interactions: Competitive locations and games." In: Francis R.L. and Iz Chaudani M.I. (eds) *Discrete location theory.* Wiley, New York, 1990, pp. 439-478.

[12] Hochbaum, D.S., A Non-Lienar Knapsack Problem, *Oper. Res. Letters*, Vol. 17, 1995, pp 103-110.

[13] Huff, D.L., "Defining and Estimating a Trade Area," *Journal of Marketing*, Vol. 28, 1964, pp 34-38.

[14] Nakanishi, M. and L.G. Cooper, "Parameter Estimates for Multiplicative Competitive Interaction Models-Least Square Approach," *Journal of Marketing Research*, Vol. 11, 1974, pp 303-311.

[15] ReVelle, C. "The maximum capture or 'sphere of influence' location problem: Hotelling revisited on a network." *Journal of Regional Science*, 26(2), 1986, pp 343-357.