Sequential Competitive Facility Location: Exact and Approximate Algorithms

Mingyao Qi
Logistics and Transportation Division, Tsinghua University Shenzhen International Graduate School, Shenzhen, China, 518055, qimy@sz.tsinghua.edu.cn

Ruiwei Jiang
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, ruiwei@umich.edu

Siqian Shen
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, siqian@umich.edu

We study a competitive facility location problem (CFLP), in which two firms sequentially select locations of new facilities, in order to maximize their market shares of customer demand that follows a probabilistic choice model. This process is a Stackelberg game and admits a bilevel mixed-integer nonlinear program (MINLP) formulation. Through integer programming methods, we derive an equivalent, single-level MINLP reformulation. In addition, we exploit the problem structures and derive two classes of valid inequalities, one based on submodularity and the other based on concave overestimation. We apply these inequalities in a branch-and-cut algorithm to find a globally optimal solution to CFLP. Furthermore, we propose an approximation algorithm for solving CFLP that is computationally more effective. Notably, this algorithm admits a constant approximation guarantee. Extensive numerical studies demonstrate that the exact algorithm can significantly accelerate the solving of CFLP on problem instances that have not been solved to optimality by existing methods. The approximation algorithm can find near-optimal solutions even more quickly.

Key words: competitive facility location; mixed-integer nonlinear programming; branch-and-cut; submodularity; concave overestimation; approximation algorithm

1. Introduction

The competitive facility location problem (CFLP) involves decision games between two or multiple firms, who compete for customer demand of substitutable products or service in a shared market. CFLP arises in a wide variety of applications including opening new retail stores, locating park-and-ride car rental facilities, building charging stations for electric vehicles, etc. In addition, it extends classical location problems, e.g., p-median and maximum coverage, to a more complex decision-making environment, in which the market no longer assumes to have a spatial monopoly but to have co-existing competitors and a certain type of consumer patronage behavior.

Competition Type: CFLP models mainly consider three types of competition: static, sequential, and dynamic (Plastria 2001). In static CFLP, a “newcomer” firm enters a market and knows a priori the existing facilities set up by its competitors, such as their locations and levels of
attractiveness. Following a certain customer behavior model which we will detail next, the firm decides where to locate new facilities to maximize its market share. Representative studies on static CFLP include Benati and Hansen (2002), Haase and Müller (2014), Ljubić and Moreno (2018), Mai and Lodi (2020) and references therein. Both sequential and dynamic CFLPs allow recourse actions of locating facilities by a firm once its competitor opens new facilities. For example, Eiselt and Laporte (1997), Plastria and Vanhaverbeke (2008), Kıcı̇kaydn et al. (2011, 2012), Kress and Pesch (2012), Drezner et al. (2015), Gentile et al. (2018) consider variants of sequential CFLP, in which a leader optimizes facility locations by taking into account a follower’s potential location choices, which are made after the leader takes actions. Therefore, one could view static CFLP as the follower’s problem in sequential CFLP, with fixed locations from the leader. In dynamic CFLP, competing firms in a market make non-cooperative decisions simultaneously and iteratively, until a Nash equilibrium, if any, is reached (Godinho and Dias 2010). In this paper, we focus on sequential CFLP, where (i) a firm can react to the competitor’s actions by locating in remaining candidate sites (as opposed to static CFLP), and (ii) location decisions are relatively expensive and long-lasting, making dynamic relocation economically prohibitive (as opposed to dynamic CFLP). A typical application of our approach is to locate chained business facilities such as hotels, supermarkets, and shopping malls, for which an investor needs to make a long-term facility deployment plan with foresight of its competitors’ response.

Customer Behavior: In a CFLP, customers are assumed to act as independent decision makers based on utilities they can receive from facilities. The utility typically depends on the distance to each facility, facility size, service price, etc. (see, e.g., O’Kelly 1999, for a related empirical study). Then, a choice behavior model can be used to translate the utilities into how customers patronize open facilities. The utility and choice model make key ingredients for CFLP and greatly affect its computational complexity as well as solution approaches. In particular, a deterministic choice model assumes that each customer purchases all the goods from a facility that has the highest utility, e.g., the closest facility. Accordingly, sequential CFLP with deterministic customer choice admits a mixed-integer linear program (MILP) formulation, which can readily be solved by commercial solvers. For related literature, we refer to Plastria and Vanhaverbeke (2008), Roboredo and Pessoa (2013), Alekseeva et al. (2015), Drezner et al. (2015), Gentile et al. (2018). In contrast, A probabilistic choice model splits customer demand across multiple facilities with certain probabilities. For example, in the well-celebrated multinomial logit (MNL) model (McFadden 1973), the probability of patronizing a facility is proportional to the natural exponential of the utility. MNL has been widely applied in static CFLP (see, e.g., Benati and Hansen 2002, Haase and Müller 2014, Ljubić and Moreno 2018, Mai and Lodi 2020), but has received much less attention in sequential CFLP (see, e.g., Kıcı̇kaydn et al. 2012), partly because it gives rise to a mixed-integer nonlinear
program (MINLP) and renders a significant computational challenge when presented in a bilevel formulation. We review details of the relevant literature in Section 2.

Focus of the Paper: We investigate the general problem of sequential CFLP with probabilistic customer choice. Our goal is to derive exact and approximate formulations, as well as solution approaches that achieve high computational efficiency for solving instances with practical sizes. In the ensuing bilevel model, both the upper- and lower-level problems are MINLPs, respectively solved by the leader and the follower. We recast the bilevel program as a single-level MINLP and derive two classes of valid inequalities to solve the reformulation exactly. In addition, we propose an alternative algorithm to solve the model approximately and show that this algorithm admits a constant approximation guarantee.

The remainder of the paper is organized as follows. In Section 2, we review prior work and position our study in the existing literature. In Section 3, we formulate the bilevel model and recast it as a single-level MINLP. Then, we derive valid inequalities in Section 4 and the approximation algorithm in Section 5. We present computational results based on instances with diverse sizes and complexity in Section 6, before we conclude and propose future research directions in Section 7.

Notation: For \( a, b \in \mathbb{R} \), we define \( a^+ := \max\{a, 0\} \) and \( a \lor b := \max\{a, b\} \). For set \( S \), \(|S|\) denotes its cardinality and \( 2^S \) denotes the collection of all its subsets. The notation \( \| \cdot \|_2 \) represents the 2-norm in the real space, and \( e \) represents an all-one vector with suitable dimension.

2. Literature Review

We focus on reviewing the most related work on static CFLP and sequential CFLP. For dynamic CFLP, we refer to Diaz-Báñez et al. (2011), Godinho and Dias (2013) for thorough reviews. Recently, Kleinert et al. (2021) survey on integer programming approaches for bilevel optimization.

Static CFLP was first proposed by Slater (1975) and was also known as the \((r \mid X_p)\)-medianoid problem, in which a decision maker locates \( r \) facilities when the \( p \) locations of its competitors’, denoted by \( X_p \), are given (see, e.g., Hakimi 1983). Plastria (2001) provided a comprehensive survey of the models and solution approaches, including heuristic methods, for static CFLP. Assuming a deterministic choice model and that both customers and facilities are located at discrete points of a network, one can model the static CFLP as a MILP (see, e.g., ReVelle 1986, ReVelle and Serra 1995). In contrast, Luce (1959) and Huff (1964) proposed a probabilistic choice model that characterizes customer behavior given utilities of multiple facilities. Benati and Hansen (2002) adopted this probabilistic choice model in static CFLP and formulated it as a MINLP. They exploited both concavity and submodularity of the objective function to construct an exact and a heuristic algorithm, respectively. In the exact algorithm, they recast this model as a MILP, which was further strengthened in Haase (2009), Aros-Vera et al. (2013), and Zhang et al. (2012) to improve
the computational performance. Recently, Haase and Müller (2014) compared the aforementioned methods via extensive computational studies and empirical analysis, and Freire et al. (2016) conducted numerical studies on diverse instances, showing that the state-of-the-art methods can handle small- to medium-sized problem instances. Ljubić and Moreno (2018) proposed a branch-and-cut algorithm using outer approximation (OA) inequalities and submodularity inequalities for solving static CFLP with random utilities. Their method outperforms the state-of-the-art approaches with two to three orders of magnitude. Mai and Lodi (2020) propose multicut OA inequities for groups of demand points and implement them in a cutting-plane approach. Different from most existing static CFLP models, Dan and Marcotte (2019) considered random utility based on both travel time and queuing delay at facilities, leading to a MINLP formulation. They derived a piecewise linear approximation and a heuristic method for solving this model. In this paper, we consider sequential CFLP, which gives rise to a more challenging, bilevel MINLP. Additionally, in terms of methodology, our valid inequalities are different from those derived in Ljubić and Moreno (2018), Mai and Lodi (2020). Specifically, Ljubić and Moreno (2018) derived OA inequalities from the concave objective function of their static CFLP model, but our sequential CFLP model undermines such concavity. Nevertheless, we are able to “bulge up” our objective function to restore concavity while retaining exactness. This yields a new class of valid inequalities that have not been developed in the existing CFLP literature and can significantly speed up the computation of the bilevel MINLP.

### Table 1: Comparison with existing work on sequential CFLP

| Reference | Choice Model | Formulation (Upper/Lower) | Solution Approach |
|-----------|--------------|---------------------------|-------------------|
| Serra and ReVelle (1994) | Deterministic | MILP | Heuristic |
| Fischer (2002) | Deterministic | Bilevel (MINLP/MILP) | Heuristic |
| Plastria and Vanhaverbeke (2008) | Deterministic | MILP | Exact; commercial solver |
| Kičińkayd et al. (2011) | Probabilistic | Bilevel (MINLP/NLP) | Exact; monolithic MINLP reformulation |
| Kičińkayd et al. (2012) | Probabilistic | Bilevel (MINLP/MINLP) | Heuristic |
| Roboredo and Pessoa (2013) | Deterministic | MILP | Exact; branch-and-cut |
| Alekseeva et al. (2015) | Deterministic | MILP | Exact; iterative method |
| Drezner et al. (2015) | Deterministic | Bilevel (MINLP/MILP) | Heuristic |
| Gentile et al. (2018) | Deterministic | MILP | Exact; branch-and-cut |
| This paper | Probabilistic | Bilevel (MINLP/MINLP) | Exact and approximate; branch-and-cut |

Sequential CFLP is also known as CFLP with leader-follower game in the literature. Accordingly, it can be modeled as a bilevel program, which involves the leader’s location model in the upper level and a static CFLP in the lower level (i.e., the follower’s location model). In Table 1, we summarize representative work of different sequential CFLP models based on deterministic/probabilistic choice models, their formulations and solution methods. In case if a bilevel formulation is adopted, we specify the formulations used in the upper/lower levels. We also indicate exact solution methods that guarantee global optimum. To the best of our knowledge, this paper provides the first exact
solution approach, as well as the first approximation algorithm with an optimality guarantee, for solving sequential CFLP with probabilistic choice model.

Assuming deterministic choice and discrete location space, one can reformulate the bilevel program of sequential CFLP as a single-level MILP (see Plastria and Vanhaverbeke 2008, Roboredo and Pessoa 2013, Alekseeva et al. 2015, Gentile et al. 2018). This MILP involves a polynomial number of variables and an exponential number of constraints. Instances with up to 100 candidate facility sites and 100 customers can be solved exactly through a branch-and-cut algorithm using the commercial solvers (see Gentile et al. 2018). Drezner and Drezner (1998) were the first to study sequential CFLP with probabilistic choice, for which they developed a heuristic algorithm (without optimality-gap guarantees). Recently, Sáiz et al. (2009) extended this work and applied a branch-and-bound algorithm to seek exact solutions. Different from Drezner and Drezner (1998) and Sáiz et al. (2009), Küğükaydn et al. (2011) optimized the leader’s facility locations with fixed locations from the follower, who only optimizes the facility attractiveness, e.g., facility sizes. In their model, the lower-level problem becomes a convex nonlinear program (NLP). It then allows the authors to use the Karush-Kuhn-Tucker (KKT) optimality conditions to seek the follower’s optimal decisions, and therefore the bilevel program becomes a single-level MINLP after adding the KKT conditions into the upper-level formulation. To the best of our knowledge, Küçükaydn et al. (2012) is the only study that considered the same sequential CFLP as ours, which assumes a probabilistic choice model and discrete location space. Küçükaydn et al. (2012) developed three heuristics for the related bilevel program with MINLP in both levels, as well as an ϵ-optimal method by iteratively fixing the leader’s decisions. Small-sized instances having up to 16 candidate facilities were solved without optimality guarantee. In this paper, we shall derive a solution approach that solves this model to global optimum on larger-sized instances (e.g., with up to 100 candidate facility sites and 2000 customers). To achieve this, we interpret sequential CFLP as a robust optimization (RO) model, in which the follower acts adversarially in order to decrease the market share of the leader’s. Nonetheless, as we shall see in Section 3, the uncertainty set of this RO model depends on the leader’s locations, leading to an intractable, decision-dependent RO model. We sidestep this challenge by deriving an equivalent, but decision-independent RO formulation. Our contribution is fourfold.

1. Without losing optimality, we revise the objective function of sequential CFLP to make the uncertainty set decision-independent. Accordingly, we recast the bilevel program as a single-level MINLP. This allows us to solve sequential CFLP to global optimum by using a finite branch-and-cut algorithm.

2. We derive two classes of valid inequalities to accelerate the branch-and-cut procedure. The first class exploits the submodularity of the revised objective function. In addition, we “bulge up” this
function to make it concave without losing any exactness. This yields a class of bulge inequalities. Inspired by RO, we further derive an approximate separation of these valid inequalities that only consumes a sorting procedure to compute.

3. We propose an approximation algorithm for solving sequential CFLP based on a mixed-integer second-order conic program (MISOCP), which can be readily solved in commercial solvers. In addition, we derive a constant approximation guarantee on the ensuing market share.

4. Through extensive computational experiments, we demonstrate that our valid inequalities can significantly accelerate the solving of sequential CFLP. For example, our approach is able to solve instances with 100 candidate facilities and 2000 customer nodes in minutes, which has never been achieved in the past (even by heuristic approaches). In addition, our approximation algorithm can obtain near-optimal solutions even more quickly.

3. Bilevel Model and Single-level Reformulation

In sequential CFLP, two firms (a leader and a follower) deploy facilities to provide substitutable commodities to customers located in a set $I$ of nodes. In each node $i \in I$, there is $h_i$ portion of the total customer demand to patronize these facilities, i.e., $\sum_{i \in I} h_i = 1$. The leader and the follower may already have existing facilities in the market, denoted by sets $J^L$ and $J^F$ respectively, with $J^L \cap J^F = \emptyset$. The new facilities may be deployed in a set $J$ of candidate sites such that $J \cap (J^L \cup J^F) = \emptyset$. The competition follows a Stackelberg game (von Stackelberg 1934), in which the leader first locates $p$ facilities to maximize her market share while foreseeing that the follower will react and locate $r$ facilities in the remaining candidate sites to maximize his own market share, where $p + r \leq |J|$. Without loss of generality, we assume that two or more facilities do not co-locate at one candidate site, because mathematically we can always split a candidate site and the corresponding utilities if needed.

We adopt the widely applied MNL model (see, e.g., McFadden 1973, Ben-Akiva et al. 1985) for probabilistic customer choice. Specifically, MNL assumes that a customer dwelling at node $i$ and patronizing a facility deployed at site $j$ receives utility $u_{ij} := \alpha_j - \beta d_{ij} + \epsilon_{ij}$, where $\epsilon_{ij}$ denotes a random noise, $-\beta < 0$ denotes the negative impact of traveling distance $d_{ij}$ between node $i$ and site $j$, and $\alpha_j$ denotes the attractiveness of facility $j$, which depends on various characteristics such as size, convenience of transportation, reputation, etc. In a seminal work, McFadden (1973) showed that, if the random noises $\epsilon_{ij}$ are independent and identically follow the standard Gumbel distribution then the probability $P_{ij}$ of the customer patronizing facility $j$ follows $P_{ij} = \frac{\exp(\alpha_j - \beta d_{ij})}{\sum_{k \in J^0} \exp(\alpha_k - \beta d_{ik})}$, where $J^0$ denotes the set of facilities deployed by either the leader or the follower. For notation brevity, we denote $w_{ij} := \exp(\alpha_j - \beta d_{ij})$ for all $i \in I$ and $j \in J \cup J^L \cup J^F$, and let $U_{i}^{L} := \sum_{j \in J^L} w_{ij}$ and $U_{i}^{F} := \sum_{j \in J^F} w_{ij}$ denote utility of the pre-existing facilities already open by the leader and the follower, respectively. For new facilities, we let binary variables $x_j$ and $y_j$, for all $j \in J$, denote if
the leader/follower deploys a facility at site $j$, respectively. Then, we calculate the leader’s market share as

$$L^+(x, y) := \sum_{i \in I} h_i \left( \frac{U_i^L + \sum_{j \in J} w_{ij} x_j}{U_i^L + U_i^F + \sum_{j \in J} w_{ij} (x_j + y_j)} \right).$$

(1)

Accordingly, we formulate sequential CFLP with probabilistic customer choice as a bilevel program:

**(S-CFLP)**

$$\max_x L^+(x, y^*)$$

s.t. $\sum_{j \in J} x_j \leq p,$

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

where $y^* \in \arg \max \sum_{i \in I} h_i \left( \frac{U_i^F + \sum_{j \in J} w_{ij} y_j}{U_i^L + U_i^F + \sum_{j \in J} w_{ij} (x_j + y_j)} \right)$

s.t. $\sum_{j \in J} y_j \leq r,$

$y_j \leq 1 - x_j, \ \forall j \in J,$

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

(2a)–(2g)

The objective function (2a) of the upper-level problem aims to maximize the leader’s market share, and constraints (2b)–(2c) ensure that the leader deploys no more than $p$ new facilities. In the lower-level problem, the objective function (2d) derives an optimal solution $y^*$ that maximizes the follower’s market share for given $x$. Constraints (2e) and (2g) ensure that the follower opens up to $r$ new facilities, and constraints (2f) prohibit co-location. Here both upper- and lower-level problems are MINLPs, giving rise to significant computational challenges. In fact, the following hardness result indicates that it is not even possible to find a good approximation of (S-CFLP) unless $P = NP$. We present a detailed proof to Theorem 1 in Appendix A.1.

**Theorem 1 (Adapted from Theorem 3 of Krause et al. (2008)).** There does not exist a polynomial-time, constant approximation algorithm for (S-CFLP) unless $P = NP$. Specifically, let $z^*$ represent the optimal value of (S-CFLP). If there exists a constant $c_0 > 0$ and an algorithm, which runs in time polynomial in $|J|, p$ and guarantee to find a solution $x$ such that $L^+(x, y^*) \geq c_0 z^*$, then $P = NP$.

Next, we take a RO perspective to recasting (S-CFLP) into a computable form. We start by noticing that the leader’s and the follower’s objective functions (2a) and (2d) sum up to 1. Intuitively, this is because a customer patronizes either the leader or the follower. This suggests that formulation (2) can be viewed as a RO model, in which the follower acts adversarially to decrease the market share of the leader’s. From this perspective, (S-CFLP) is equivalent to a RO model:

$$\max_{x \in X} \min_{y \in \mathcal{Y}(x)} L^+(x, y),$$

(3)
where $\mathcal{X} := \{x \in \{0,1\}^{|J|} : e^T x \leq p\}$ denotes the leader’s feasible region and $\mathcal{Y}(x) := \{y \in \{0,1\}^{|J|} : (2e)−(2g)\}$ denotes the follower’s, which is interpreted as an uncertainty set in RO. Unfortunately, this uncertainty set is decision-dependent, preventing us from applying standard reformulation techniques. For example, one might suggest to solve formulation (3) by rewriting the inner formulation in a hypergraphic form and applying delayed constraint generation (DCG). Specifically, one can rewrite (3) as $\max_{x \in \mathcal{X}, \theta^+} \{\theta^+: \theta^+ \leq L^+(x,y), \forall y \in \mathcal{Y}(x)\}$ and iteratively incorporates inequalities (cuts) $\theta^+ \leq L^+(x,y)$ only when they are violated. Unfortunately, this does not apply to (3) either, because $\mathcal{Y}(x)$ involves binary restrictions (2g) and hence strong duality fails to hold for the inner formulation. Note that constraints (2g) may not be naïvely relaxed because $L^+(x,y)$ is convex in $y$. For the same reason, the exactness of this formulation would be lost if we replace the inner formulation with its KKT conditions.

To make the uncertainty set decision-independent, we revise the objective function $L^+(x,y)$ so that co-location constraints (2f) can be relaxed without any loss of optimality. This then allows us to apply DCG. We specify the result in the following theorem.

**Theorem 2.** Define $\theta^+, \theta : \{0,1\}^{|J|} \rightarrow \mathbb{R}$ such that $\theta^+(x) := \min_{y \in \mathcal{Y}(x)} L^+(x,y)$ and $\theta(x) := \min_{y \in \mathcal{Y}} L(x,y)$, where $\mathcal{Y} := \{y : (2e), (2g)\}$ and

$$L(x,y) := \sum_{i \in I} h_i \left( \frac{U_i^L + \sum_{j \in J} w_{ij} x_j}{U_i^L + U_i^F + \sum_{j \in J} w_{ij} (x_j \lor y_j)} \right).$$

(4)

Then, it holds that $\theta^+(x) = \theta(x)$ for all $x \in \mathcal{X}$.

**Proof of Theorem 2:** Define $F_0 := \mathcal{Y}$ and $F_x := \{y \in \{0,1\}^{|J|} : y_j \leq 1−x_j, \forall j \in J\}$. Then, for any $x \in \mathcal{X}$, we have $\theta^+(x) = \min_{y \in F_0 \cap F_x} L^+(x,y)$ and $\theta(x) = \min_{y \in \mathcal{Y}} \{\theta_1(x), \theta_2(x)\}$, where $\theta_1(x) = \min_{y \in F_0 \cap F_x} L(x,y)$ and $\theta_2(x) = \min_{y \in F_0 \cap F_x} L(x,y)$. In addition, for any $y \in F_0 \cap F_x$, we have $L^+(x,y) = L(x,y)$ because $x_j + y_j = x_j \lor y_j$ for all $j \in J$. Hence, $\theta^+(x) = \theta_1(x)$. It remains to show that $\theta_1(x) \leq \theta_2(x)$ for all $x \in \mathcal{X}$ and then the equivalence can be drawn between $\theta^+(x)$ and $\theta(x)$.

To this end, for any $x \in \mathcal{X}$ and $\bar{y} \in F_0 \cap F_x$, we construct a $\tilde{y} \in F_0 \cap F_x$ such that $L(x, \tilde{y}) \leq L(x, \bar{y})$. Since $\bar{y} \in F_0 \cap F_x$, there exists a nonempty subset $K \subseteq J$ such that (i) $\bar{y}_j > 1−x_j$, i.e., $x_j = \bar{y}_j = 1$ for all $j \in K$ and (ii) $\bar{y}_j \leq 1−x_j$ for all $j \in J \setminus K$. We claim that there exists a subset $M \subseteq J \setminus K$
with $|M| = |K|$ and $x_j = \overline{y}_j = 0$ for all $j \in M$. To see this, we denote $J_{mn} := \{j \in J : x_j = m, \overline{y}_j = n\}$ for $m, n \in \{0, 1\}$. Then, it holds that
\[
|J_{00}| = |J| - |J_{11}| - |J_{01}| - |J_{10}|
\geq |J| - |K| - (r - |K|) - (p - |K|)
= |K| + (|J| - r - p) \geq |K|
\]
where the first inequality is because $\sum_{j \in J} \overline{y}_j \leq r$ and $\sum_{j \in J} x_j \leq p$ and the last inequality holds because $p + r \leq |J|$. Then, the existence of $M$ follows from the pigeonhole principle. Now define a $\tilde{y} \in \{0, 1\}^{|J|}$ such that $\tilde{y}_j = \overline{y}_j$ for all $j \in J \setminus (K \cup M)$, $\tilde{y}_j = 0$ for all $j \in K$, and $\tilde{y}_j = 1$ for all $j \in M$. Then, $\tilde{y} \in F_0 \cap F_2$ by construction. In addition, for each $i \in I$, we have
\[
\sum_{j \in J} w_{ij}(x_j \vee \overline{y}_j)
= \sum_{j \in J} w_{ij}(x_j \vee \overline{y}_j) + \sum_{j \in M} w_{ij}(x_j \vee \overline{y}_j) + \sum_{j \in J \setminus (K \cup M)} w_{ij}(x_j \vee \overline{y}_j)
= \sum_{j \in K} w_{ij} \times 1 + \sum_{j \in M} w_{ij} \times 0 + \sum_{j \in J \setminus (K \cup M)} w_{ij}(x_j \vee \overline{y}_j)
\leq \sum_{j \in K} w_{ij} \times 1 + \sum_{j \in M} w_{ij} \times 1 + \sum_{j \in J \setminus (K \cup M)} w_{ij}(x_j \vee \overline{y}_j)
= \sum_{j \in K} w_{ij}(x_j \vee \tilde{y}_j) + \sum_{j \in M} w_{ij}(x_j \vee \tilde{y}_j) + \sum_{j \in J \setminus (K \cup M)} w_{ij}(x_j \vee \overline{y}_j)
= \sum_{j \in J} w_{ij}(x_j \vee \tilde{y}_j).
\]
As a result, $L(x, \overline{y}) \geq L(x, \tilde{y})$ and the proof is completed. □

Theorem 2 recasts the bilevel model (S-CFLP) as the following single-level MINLP:
\[
\begin{align*}
\max_{x \in X, \theta} & \quad \theta \\
\text{s.t.} & \quad \theta \leq L(x, y), \quad \forall y \in \mathcal{Y}. & \quad (5a)
\end{align*}
\]
Although model (5) incorporates an exponential number of constraints due to the cardinality of set $\mathcal{Y}$, it can be solved through DCG. Specifically, we relax constraints (5b) and iteratively add them back if needed. In each iteration, we obtain an incumbent solution $(\hat{x}, \hat{\theta})$ from the relaxed formulation. Then, we solve the following separation problem
\[
\min_y \left\{ L(\hat{x}, y) : y \in \mathcal{Y} \right\} & \quad (6)
\]
to decide if this solution violates any of constraints (5b). If not, then $(\hat{x}, \hat{\theta})$ is an optimal solution to (5); otherwise, we find a $\hat{y} \in \mathcal{Y}$ such that constraint $\theta \leq L(x, \hat{y})$ is violated, i.e., $\hat{\theta} > L(\hat{x}, \hat{y})$. We append this violated constraint in the iteratively-solved relaxed formulation of (5) to cut off the incumbent solution. Since $\mathcal{Y}$ is finite, DCG terminates with a global optimal solution in a finite number of iterations. We summarize the full details of our approach in Algorithm 1 at the end of Section 4, after completing the derivation of the valid inequalities and the approximate separation.
4. Valid Inequalities and Approximate Separation

There are two challenges on applying DCG. First, the violated constraints we incorporate, $\theta \leq L(x, \hat{y})$, are nonlinear. As a consequence, in every iteration we need to solve the relaxed formulation of model (5) as a MINLP. Moreover, as we shall see in Section 4.2, function $L(x, \hat{y})$ is non-concave in $x$ in its current form presented in (4). That is, the relaxed formulation remains a non-convex NLP even if we further relax its integer restrictions. To address these challenges, we derive two classes of linear valid inequalities in Sections 4.1 and 4.2, respectively. Together, they generate a tight MILP relaxation of the nonlinear, non-convex formulation, which can be readily solved by off-the-shelf solvers. Second, the separation problem (6) by itself is a MINLP. To address this, we recall that (6) is equivalent to a static CFLP model and solve it via the state-of-the-art approach from Ljubić and Moreno (2018). As a further improvement, in Section 4.3, we derive an approximate, but much faster, approach to solving (6) via a single-round sorting. In Section 6, we shall demonstrate that our approximate separation yields a substantial speedup through extensive computational studies.

4.1. Submodular Inequalities

We start by recalling the following definition of submodular functions.

**Definition 1 (Submodular Functions).** A function $f : 2^J \rightarrow \mathbb{R}$ is submodular if

$$f(S \cup \{j\}) - f(S) \geq f(R \cup \{j\}) - f(R)$$

for all subsets $S \subseteq R \subseteq J$ and all element $j \in J \setminus R$. □

Intuitively, $f$ is submodular if the marginal gain of incorporating any additional element $j$ is non-increasing in the subset $S$. We show that, for any fixed $y \in \mathcal{Y}$, the function $L(x, y)$ is submodular with respect to the index set $X$ of $x$, i.e., $X := \{j \in J : x_j = 1\}$. This observation enables us to represent the nonlinear constraint $\theta \leq L(x, y)$ as a set of linear inequalities.

To state the results formally, we denote $Y := \{j \in J : y_j = 1\}$ as the index set of $y$. In addition, we define set functions $L_Y : 2^J \rightarrow \mathbb{R}_+$ and $f_{i,Y} : 2^J \rightarrow \mathbb{R}_+$ such that $L_Y(X) := \sum_{i \in I} h_i f_{i,Y}(X)$ and

$$f_{i,Y}(X) := \frac{U_i^L + \sum_{j \in X} w_{ij}}{U_i^L + U_i^Y + \sum_{j \in X \cup Y} w_{ij}}$$

for all $i \in I$. Intuitively, $f_{i,Y}(X)$ evaluates the percentage of demand from customer $i$ that patronizes the leader’s facilities and $L_Y$ evaluates the total market share of the leader, if the leader and the follower deploy facilities in sets $X$ and $Y$, respectively. Hence, it holds that $L(x, y) = L_Y(X)$. Now we formally show the submodularity of function $L_Y$ and present a proof in Appendix A.2.

**Proposition 1.** For any $Y \subseteq J$, $L_Y$ is submodular.
Since $L_Y$ is submodular, we follow Nemhauser and Wolsey (1981) to rewrite the constraint $\theta \leq L(x, y) \equiv L_Y(X)$ as a set of linear inequalities.

**Proposition 2 (Adapted from Theorem 6 of Nemhauser and Wolsey (1981)).** For any $y \in Y$, the constraint $\theta \leq L(x, y)$ is equivalent to the following linear constraints:

$$\theta \leq L_Y(S) - \sum_{k \in S} \rho_Y(J \setminus \{k\}; k)(1 - x_k) + \sum_{k \in J \setminus S} \rho_Y(S; k)x_k, \quad \forall S \subseteq J,$$

where $\rho_Y(S; k) := L_Y(S \cup k) - L_Y(S)$ for all $S \subseteq J$ and $k \in J \setminus S$.

Constraints (8) involve an exponential number of inequalities. Hence, in DCG, a straightforward replacement of $\theta \leq L(x, y)$ with (8) drastically increases the formulation size. Instead, we can replace $\theta \leq L(x, y)$ with the most violated inequality among (8), which serves the same purpose of cutting off the incumbent solution. Specifically, for given $(\hat{x}, \hat{\theta})$, inequalities (8) hold valid iff

$$\hat{\theta} \leq \min_{S \subseteq J} \left\{ L_Y(S) - \sum_{k \in S} \rho_Y(J \setminus \{k\}; k)(1 - \hat{x}_k) + \sum_{k \in J \setminus S} \rho_Y(S; k)\hat{x}_k \right\},$$

and to find the most violated inequality it suffices to solve the combinatorial optimization problem on the right-hand side of (9). In what follows, we show that this task can be accomplished efficiently, in polynomial time.

Suppose that $\hat{x} \in \{0, 1\}^{|J|}$. This can take place when a relaxation of model (5) happens to produce a binary-valued solution or at a leaf node of the branch-and-bound tree for solving (5). For this case, Ljubić and Moreno (2018) showed that the index set $\hat{X}$ of $\hat{x}$, i.e., $\hat{X} := \{j \in J : \hat{x}_j = 1\}$ is optimal to problem (9). That is, the most violated inequality among (8) is the one with $S = \hat{X}$.

More generally, we consider $\hat{x} \in [0, 1]^{|J|}$, i.e., $\hat{x}$ can be either fractional- or binary-valued. This can take place at any node of the branch-and-bound tree, where we relax (a part of or all of) the binary restrictions on variables $x$. The next proposition shows that problem (9) has a submodular objective function and so it admits a polynomial-time solution (see Edmonds 1970, Topkis 1978).

**Proposition 3.** For any $x \in [0, 1]^{|J|}$ and $Y \subseteq J$, define $H : 2^J \rightarrow \mathbb{R}$ such that $H(S) := L_Y(S) - \sum_{k \in S} \rho_Y(J \setminus k; k)(1 - x_k) + \sum_{k \in J \setminus S} \rho_Y(S; k)x_k$. Then, $H$ is submodular.

The detailed proof is presented in Appendix A.3. Proposition 3 indicates that we can find the most violated inequality among (8) in polynomial time. When implementing DCG in commercial solvers (e.g., CPLEX or GUROBI), we can incorporate the most violated inequalities via the lazy callback when $\hat{x}$ are binary-valued or via the user callback when $\hat{x}$ are fractional-valued.
4.2. Bulge Inequalities

A second alternative of the nonlinear constraints $\theta \leq L(x,y)$ is the supporting hyperplanes of the hypograph of function $L(x,y)$, also known as the outer approximation method (see, e.g., Duran and Grossmann 1986, Ljubić and Moreno 2018). To this end, we extend the domain of $L(x,y)$ by defining

$$\hat{L}(x,y) := \sum_{i \in I} h_i \left( \frac{U_i^t + \sum_{j \in J} w_{ij} x_j}{U_i^t + U_i^p + \sum_{j \in J} w_{ij} [(1-y_j)x_j + y_j]} \right),$$

where we replace the $x_j \vee y_j$ in $L(x,y)$ with $(1-y_j)x_j + y_j$. Note that $L(x,y)$ coincides with $\hat{L}(x,y)$ whenever $(x,y)$ are binary-valued, but $\hat{L}(x,y)$ is well-defined on $[0,1]^{|J|}$. For fixed $y \in Y$, we can replace $\theta \leq L(x,y)$ with a supporting hyperplane if $\hat{L}(x,y)$ is concave in $x$. Unfortunately, this fails to hold as evidenced by the following example.

**Example 1 (Non-concavity of $\hat{L}$).** Suppose that $I = \{1\}$ and $h_1 = 1$, i.e., there is one single customer node. In this example, we ignore the index $i$ for notation simplicity. In addition, suppose that $J = \{1,2\}$, $U_i^t = U_i^p = 0$, $w_1 = w_2 = 1$, and $y_1 = 1 - y_2 = 0$. Then, $\hat{L}(x,y) = \frac{x_1 + x_2}{x_1 + 1}$ and its Hessian reads

$$-\frac{1}{(x_1 + 1)^2} \begin{bmatrix} \frac{2(1-x_2)}{x_1 + 1} & 1 \\ 1 & 0 \end{bmatrix},$$

which is not negative semidefinite on $(0,1)^{|J|}$. Therefore, $\hat{L}(x,y)$ is not concave in $x$. In particular, restricting $\hat{L}(x,y)$ on the line $x_1 + x_2 = 1$ yields $\frac{1}{x_1 + 1}$, which is in fact a convex function. $\square$

The above example suggests that we should “bulge up” $\hat{L}(x,y)$ in order to obtain a concave hypograph. To this end, we replace the linear term $x_j$ in the numerator of the $\hat{L}(x,y)$ definition with a larger, quadratic term. This yields a desired concave function as shown next.

**Proposition 4.** For fixed $y \in \{0,1\}^{|J|}$, define $\tilde{L} : \{0,1\}^{|J|} \to \mathbb{R}_+$ such that

$$\tilde{L}(x,y) := \sum_{i \in I} h_i \left( \frac{U_i^t + \sum_{j \in J} w_{ij} [-y_j x_j^2 + (1+y_j)x_j]}{U_i^t + U_i^p + \sum_{j \in J} w_{ij} [(1-y_j)x_j + y_j]} \right).$$

(10)

Then, $\tilde{L}(x,y)$ is concave in $x$. In addition, $\tilde{L}(x,y) = L(x,y)$ for all $x \in \{0,1\}^{|J|}$.

**Proof of Proposition 4:** For all $x,y \in \{0,1\}^{|J|}$, it is easy to verify $x_j \vee y_j = (1-y_j)x_j + y_j$ and $x_j = -y_j x_j^2 + (1+y_j)x_j$. It follows that $\tilde{L}(x,y) = L(x,y)$ for all $x \in \{0,1\}^{|J|}$.

It remains to show the concavity of $\tilde{L}(x,y)$. Since the sum of concave functions is concave, it suffices to show that $\tilde{L}_i(x) := \frac{U_i^t + \sum_{j \in J} w_{ij} (-y_j x_j^2 + (1+y_j)x_j)}{U_i^t + U_i^p + \sum_{j \in J} w_{ij} [(1-y_j)x_j + y_j]}$ is concave for all $i \in I$. For ease of exposition, we denote its numerator $Q := U_i^t + \sum_{j \in J} w_{ij} [-y_j x_j^2 + (1+y_j)x_j]$, its denominator $P := U_i^t + U_i^p + \sum_{j \in J} w_{ij} [(1-y_j)x_j + y_j]$, and its Hessian $H := [h_{kl}]$. It follows that

$$\frac{\partial \tilde{L}_i(x)}{\partial x_k} = \frac{-w_{ik}(1-y_k)Q}{P^2} + \frac{w_{ik}(-2y_k x_k + 1 + y_k)}{P}, \quad \forall k \in J,$$
If $y \in J, y \notin J$ then $h_{kk} = -\frac{2}{P} w_{ik}$ and $h_{kl} = 0.$

Case 2. If $y \in J_0, y \notin J_1$ then $h_{kk} = -\frac{2w_i^2}{P^2} (P - Q)$ and $h_{kl} = -\frac{2w_i w_{ij}}{P^2} (1 - x_k).$

Case 3. If $y \in J_1, y \notin J_0$ then $h_{kk} = -\frac{2w_i^2}{P^2} (P - Q)$ and $h_{kl} = -\frac{2w_i w_{ij}}{P^2} (1 - x_k).$

Case 4. If $y \in J_0, y \notin J_1$ then $h_{kk} = -\frac{2w_i^2}{P^2} (P - Q)$ and $h_{kl} = -\frac{2w_i w_{ij}}{P^2} (1 - x_k).$

Denote $J_1 := \{j \in J : y_j = 1\}$ and $J_0 := \{j \in J : y_j = 0\}.$ We simplify the expression of $H$ by examining the following four cases:

We notice that

$$P - Q = U_i^p + \sum_{j \in J} w_{ij} \left\{ \left[ (1 - y_j)x_j + y_j \right] - \left[ -y_j x_j^2 + (1 + y_j) x_j \right] \right\}$$

$$= U_i^p + \sum_{j \in J} w_{ij} (1 - x_j)^2 = U_i^p + \sum_{j \in J_1} w_{ij} (1 - x_j)^2. \quad (12)$$

Plugging (12) into (11) yields

$$v^T Hv = -\frac{2}{P^3} \left( \sum_{k \in J_1} (\sqrt{w_{ik}}(1 - x_k))^2 \sum_{k \in J_0} (w_{ik} v_k)^2 + \sum_{k \in J_1} \left( P \sqrt{w_{ik}} v_k \right)^2 \right).$$
\[ \begin{align*}
+2\left( \sum_{k \in J_1} P(1 - x_k)w_{ik}v_k \right) & \sum_{\ell \in J_0} w_{i\ell}v_{\ell} + U_i^P \left( \sum_{k \in J_0} w_{ik}v_k \right)^2 \right) \\
& = -\frac{2}{P^3} \sum_{k \in J_1} \left( \sqrt{w_{ik}(1 - x_k)} \sum_{\ell \in J_0} w_{i\ell}v_{\ell} \right)^2 + \left( P\sqrt{w_{ik}v_k} \right)^2 + 2P(1 - x_k)w_{ik}v_k \sum_{\ell \in J_0} w_{i\ell}v_{\ell} \\
& - \frac{2U_i^P}{P^3} \left( \sum_{k \in J_0} w_{ik}v_k \right)^2 \\
& = -\frac{2}{P^3} \sum_{k \in J_1} \left( \sqrt{w_{ik}(1 - x_k)} \sum_{\ell \in J_0} w_{i\ell}v_{\ell} + P\sqrt{w_{ik}v_k} \right)^2 - \frac{2U_i^P}{P^3} \left( \sum_{k \in J_0} w_{ik}v_k \right)^2 \leq 0.
\end{align*} \]

This completes the proof. \(\Box\)

![Figure 1](image)

**Figure 1** Illustration of functions \(\tilde{L}\) and \(\hat{L}\) with \(y = [0, 1]^T\)

Proposition 4 indicates that \(\hat{L}(x, y)\) is a concave representation of function \(L(x, y)\). That is, \(\hat{L}\) bulge up \(L\) to make it concave while retaining the exactness at any binary-valued \(x\). We illustrate this observation below.

**Example 2 (Concavity of \(\hat{L}\)).** Continuing from Example 1, we construct \(\hat{L}(x, y) = \frac{x_1 - x_2^2 + 2x_2}{x_1 + 1}\) for the same \(y = [0, 1]^T\). Then, its Hessian reads

\[
-\frac{2}{(x_1 + 1)^3} \begin{bmatrix} (1-x_2)^2 & (x_1+1)(1-x_2)^2 \\ (x_1+1)(1-x_2)^2 & (x_1+1) \end{bmatrix} = -\frac{2}{(x_1 + 1)^3} \begin{bmatrix} 1-x_2 & 1-x_2 \\ x_1+1 & x_1+1 \end{bmatrix}^T,
\]

which is negative semidefinite. Therefore, \(\hat{L}(x, y)\) is concave in \(x\). In particular, restricting \(\hat{L}(x, y)\) on the line \(x_1 + x_2 = 1\) yields \(3 - (x_1 + 1) - \frac{1}{x_1 + 1}\), which is a concave function. In Fig. 1, we depict functions \(\tilde{L}\) and \(\hat{L}\), as well as their restrictions on the line \(x_1 + x_2 = 1\). \(\Box\)
Thanks to the concave representation, we can replace the non-convex constraints \( \theta \leq \widehat{L}(x, y) \) with the convex ones \( \theta \leq \overline{L}(x, y) \) when solving model (5) in DCG. As a double-check, we show that the latter ones are in fact second-order conic representable and present a proof in Appendix A.4.

**Proposition 5.** For fixed \( y \in \{0, 1\}^{|J|} \), define \( J_0 := \{ j \in J : y_j = 0 \} \), \( J_1 := \{ j \in J : y_j = 1 \} \), and \( s_i(y) := U_i^L + U_i^F + \sum_{j \in J} w_{ij}y_j \) for all \( i \in I \). Then, the inequality \( \theta \leq \widehat{L}(x, y) \) holds valid if and only if there exist \( \{ \theta_i \}_{i \in I} \) such that \( \theta \leq \sum_{i \in I} h_i \theta_i \) and

\[
\left\| \sum_{j \in J} w_{ij}(1 - y_j)x_j + s_i(y) + \theta_i - 1 \right\|_2 \leq \sum_{j \in J} w_{ij}(1 - y_j)x_j + s_i(y) - \theta_i + 1, \quad \forall i \in I. \tag{13}
\]

In implementation, we replace constraints \( \theta \leq L(x, y) \) with a supporting hyperplane of \( \widehat{L}(x, y) \). Specifically, for given \((\hat{x}, \hat{y})\) in DCG, we incorporate the linear inequality

\[
\theta \leq \widehat{L}(\hat{x}, \hat{y}) + \sum_{j \in J} g_j(\hat{x}, \hat{y})(x_j - \hat{x}_j), \tag{14}
\]

where, for all \( j \in J \),

\[
g_j(\hat{x}, \hat{y}) := \frac{\partial \widehat{L}(x, y)}{\partial x_j} \bigg|_{x = \hat{x}} = \sum_{i \in I} h_i \left( \frac{-w_{ik}(1 - \hat{y}_k)Q}{P^2} + \frac{w_{ik}(-2\hat{y}_k\hat{x}_k + 1 + \hat{y}_k)}{P} \right),
\]

\[
P = U_i^L + U_i^F + \sum_{j \in J} w_{ij}[(1 - \hat{y}_j)\hat{x}_j + \hat{y}_j], \quad \text{and} \quad Q = U_i^L + \sum_{j \in J} w_{ij}[-\hat{y}_j\hat{x}_j^2 + (1 + \hat{y}_j)\hat{x}_j].
\]

In what follows, we call (14) the bulge inequalities and incorporate them via the lazy callback.

### 4.3. Approximate Separation

Here we derive an approximate approach to solving the separation problem (6). Although this problem can be seen as a static CFLP model, for which Ljubić and Moreno (2018) has provided an exact solution approach, our goal is to solve it significantly faster in order to accelerate DCG. We demonstrate the resulting speedup numerically in Section 6. We start by identifying the following optimality conditions for formulations (S-CFLP) and (6), respectively.

**Lemma 1.** There exists an optimal solution \( x^* \) to (S-CFLP) such that \( e^T x^* = p \). In addition, for any fixed \( \hat{x} \in X \), there exists an optimal solution \( y^* \) to the separation problem (6) such that \( e^T y^* = r \).

We present a proof in Appendix A.5. Following Lemma 1, we rewrite the separation problem (6) as \( \min_{y \in \overline{Y}} L(\hat{x}, y) \) for given \( \hat{x} \in X \), where \( \overline{Y} := \{ y \in \{0, 1\}^J : e^T y = r \} \). In what follows, we approximate the nonlinear function \( L(\hat{x}, y) \) from above by using a linear one. This leads to a relaxation of (6) that can be solved by a single-round sorting. The proof of the next proposition is presented in Appendix A.6.
Proposition 6. For fixed \( \hat{x} \in [0, 1]|J| \), define constants \( a_i(\hat{x}) := U_i^f + \sum_{j \in J} w_{ij}(1 - \hat{x})y_j \), and \( w_i^r(\hat{x}) := \max_{y \in \mathcal{Y}} \left\{ U_i^r + \sum_{j \in J} w_{ij}(1 - \hat{x})y_j \right\} \) for all \( i \in I \). In addition, define

\[
\alpha(\hat{x}) := \sum_{i \in I} h_i \left[ \frac{a_i(\hat{x}) (a_i(\hat{x}) + w_i^r(\hat{x}) + w_i^r(\hat{x}) - U_i^r)}{(a_i(\hat{x}) + w_i^r(\hat{x}))(a_i(\hat{x}) + w_i^r(\hat{x}))} \right],
\]

and vector \( \beta(\hat{x}) := [\beta_1(\hat{x}), \ldots, \beta_{|J|}(\hat{x})]^T \) with

\[
\beta_j(\hat{x}) := \sum_{i \in I} h_i \left[ \frac{a_i(\hat{x})w_{ij}(1 - \hat{x})}{(a_i(\hat{x}) + w_i^r(\hat{x}))(a_i(\hat{x}) + w_i^r(\hat{x}))} \right]
\]

for all \( j \in J \). Then, it holds that

\[
L(\hat{x}, y) \leq \alpha(\hat{x}) - \beta(\hat{x})^T y, \quad \forall y \in \mathcal{Y}.
\]

In addition, the problem \( \min_{y \in \mathcal{Y}} \left\{ \alpha(\hat{x}) - \beta(\hat{x})^T y \right\} \) admits an optimal solution \( \hat{y} = \{[1], \ldots, [r]\} \), where \( \{[j] : j \in J\} \) is a sorting of the set \( J \) such that \( \beta_{|J|}(\hat{x}) \geq \beta_{|J|}(\hat{x}) \).

We notice that, for given \( \hat{x} \in [0, 1]|J| \), the values of \( \alpha(\hat{x}) \) and \( \{\beta_j(\hat{x}) : j \in J\} \) can all be computed in closed-form. In addition, Proposition 6 suggests an approximate separation. Specifically, suppose that an incumbent solution \( (\hat{x}, \hat{\theta}) \), obtained from solving a relaxed formulation of (5), is such that \( \hat{\theta} > \alpha(\hat{x}) - \beta(\hat{x})^T \hat{y} \), where \( \hat{y} \) is described in Proposition 6. Then, this solution violates the constraint \( \theta \leq L(x, \hat{y}) \), which can be added back to the relaxed formulation of (5). This approximate separation reduces the effort of solving an MINLP to a single-round sorting. As we shall report in Section 6, this leads to a substantial speedup.

4.4. A Branch-and-Cut Framework

In Algorithm 1, we summarize the DCG algorithm for solving (S-CFLP), or equivalently its reformulation (5), in a branch-and-cut framework. Since there are a finite number of cuts \( \theta \leq L(x, \hat{y}) \) to add and a finite number of candidate solutions of \( \hat{x} \), we make the following claim.

Proposition 7. Algorithm 1 terminates in a finite number of steps with a global optimal solution to (S-CFLP).

5. Approximation Algorithm

In view of the computational challenges of solving (S-CFLP) to global optimum, we propose an approximate algorithm to obtain a near-optimal solution, but faster. Specifically, instead of solving a bilevel MINLP, the proposed algorithm solves a single-level MISOCPS, which can be readily solved by off-the-shelf solvers directly, waiving the need to design specialized solution methods like Algorithm 1. Notably, this algorithm also admits a constant approximation guarantee, as detailed in the following theorem. We present a proof in Appendix A.7.
Algorithm 1: A Branch-and-Cut Framework for Solving (S-CFLP)

1. **Initialization**: create a queue for formulations and insert the formulation
   \[
   \max_{x, \theta} \{ \theta : e^\top x = p, x \in [0, 1]^{J}, \theta \in [0, 1] \}, \]
   a continuous relaxation of (5), into the queue;

2. Set \( \text{BestSol} \leftarrow \emptyset \) and \( \theta_{LB} \leftarrow 0 \);

3. while the queue is non-empty do
   4. Remove a formulation from the queue;
   5. Solve the formulation and obtain an incumbent solution \( (\hat{x}, \hat{\theta}) \) with optimal value \( \hat{\theta} \);
   6. Solve either the approximate separation problem \( \min_{y \in \mathcal{Y}} \{ \alpha(\hat{x}) - \beta(\hat{x})^\top y \} \) or the exact separation problem (6) and obtain an optimal solution \( \hat{y} \);
   7. if \( \hat{\theta} > L(\hat{x}, \hat{y}) \) then
      8. Add a cut \( \theta \leq L(x, y) \) based on either the submodular inequality (8) or the bulge inequality (14) to strengthen the formulation;
      9. Insert the strengthened formulation back into the queue;
   10. else if \( \hat{\theta} > \theta_{LB} \) and \( \hat{x} \) is integral then
      11. Set \( \text{BestSol} \leftarrow \hat{x} \) and \( \theta_{LB} \leftarrow \hat{\theta} \);
   12. else if \( \hat{\theta} > \theta_{LB} \) and \( \hat{x} \) is fractional then
      13. Branch on \( \hat{x} \) and insert the two resulting formulations into the queue;

Theorem 3. Let \( z^* \) represent the optimal value of (S-CFLP), \( x^H \) represent an optimal location decision to the following MISOCP

\[
\min_{x \in X, \mu \geq 0, \nu \geq 0} \sum_{i \in I} h_i U_i^p s_i + r \lambda + e^\top \mu
\]

s.t.

\[
\lambda + \nu_j \geq \sum_{i \in I} h_i w_{ij} t_{ij}, \quad \forall j \in J,
\]

\[
2 \left[ U_i^L + \sum_{k \in J} w_{ik} x_k - s_i \right] \leq s_i + U_i^L + \sum_{k \in I} w_{ik} x_k, \quad \forall i \in I,
\]

\[
2(1 - x_j) \left[ U_i^L + \sum_{k \in J} w_{ik} x_k - t_{ij} \right] \leq t_{ij} + U_i^L + \sum_{k \in J} w_{ik} x_k, \quad \forall i \in I, j \in J.
\]

and \( z^H \) represents the objective function value of \( x^H \) in (S-CFLP), i.e., \( z^H := \min_{y \in \mathcal{Y}} L(x^H, y) \). Then, it holds that

\[
\frac{4 \gamma_\mathcal{M} \gamma_m}{(\gamma_\mathcal{M} + \gamma_m)^2} \leq \frac{z^H}{z^*} \leq 1,
\]

where

\[
\gamma_m := \min_{i \in I} \left\{ \frac{1}{1 + \max_{y \in \mathcal{Y}} \{ U_i^L + \sum_{j \in J} w_{ij} y_j \}} \right\}, \quad \gamma_\mathcal{M} := \max_{i \in I} \left\{ \frac{1}{1 + \min_{x \in \mathcal{X}} \{ U_i^L + \sum_{j \in J} w_{ij} x_j \}} \right\}
\]

with \( \mathcal{X} := \{ x \in \{0, 1\}^{|J|} : e^\top x = p \} \) and \( \mathcal{Y} := \{ y \in \{0, 1\}^{|J|} : e^\top y = r \} \).
Remark 1. We notice that constants $\gamma_M$ and $\gamma_m$ can be computed in closed-form. Hence, Theorem 3 presents a constant approximation algorithm for (S-CFLP) via solving a MISOCP. In theory, it is impossible to improve this result (i.e., to a polynomial-time approximation algorithm) unless $P = NP$, in view of the inapproximability conclusion of Theorem 1. Nevertheless, The MISOCP is practically tractable thanks to promising performance of state-of-the-art solvers. In computation, we can further improve the efficacy of solving formulation (15) by adding valid inequalities similar to those presented in Section 4. We present the details of these inequalities in Appendix B.

6. Computational Results

We test a variety of sequential CFLP instances to validate the efficacy of our approaches and analyze the numerical results. We describe configurations and instance design in Section 6.1 and demonstrate that

(i) the submodular inequalities (8) and bulge inequalities (14) significantly accelerate the branch-and-cut framework for solving (S-CFLP) in Section 6.2;
(ii) the approximate separation provides a further speed-up in Section 6.3;
(iii) the approximation algorithm can quickly find a near-optimal solution in Section 6.4;
(iv) the pattern of (leader’s and follower’s) optimal locations highly depends on the customer choice model in Section 6.5, generating insights on winning market share.

6.1. Configurations and Instance Design

We solve all mixed-integer programs in CPLEX 12.6 using the default configurations. All the codes are written in C++. All computations are run on a PC with Intel CORE (TM) i7-8550 1.8GHz CPU, 16G RAM running 64-bit Windows 10.

To the best of our knowledge, there are no immediately available benchmark instances having comparable sizes to what we aim to solve in the existing literature. Therefore, we randomly generate our instances following the settings in the sequential CFLP literature including Küçükaydın et al. (2012), Haase and Müller (2014), Gentile et al. (2018). In specific, we consider a $[0, 50] \times [0, 50]$ square on a planar surface, in which the locations of demand points and candidate facility sites are randomly generated with integer coordinates. We use the Euclidean metric to compute distances and a MNL model with $\beta := 0.1$ to generate probabilistic utilities. In Section 6.5, we conduct sensitivity analysis on the value of $\beta$. Without loss of generality, we set $\alpha_j := 0$ for all $j \in J$ to indicate identical attractiveness of each facility. We also assume that $J^L = J^F = \emptyset$.

Before introducing our results, we briefly review the scale of experiments conducted by previous studies on sequential CFLP in Table 2. Note that except for Küçükaydın et al. (2011), where five candidate sites are considered, none of the previous studies can guarantee solution optimum. In this paper, we challenge to solve problems with up to 100 facility sites and 2000 customer nodes to global optimum.
Table 2  Parameter settings and instant sizes in state-of-the-art literature

| Reference                  | $|I|$ | $|J|$ | $p$ | $r$ |
|----------------------------|------|------|-----|-----|
| Kılıçkaydn et al. (2011)   | 30   | 5    | *   | 0   |
| Kılıçkaydn et al. (2012)   | 100  | 20   | *   | *   |
| Roboredo and Pessoa (2013) | 100  | 50   | 4   | 4   |
| Alekseeva et al. (2015)    | 100  | 100  | 20  | 20  |
| Gentile et al. (2018)      | 225  | 225  | 5   | 5   |

Table 3  Performance of Algorithm 1

| Instance     | SC | BI | SCBI | Benchmark |
|--------------|----|----|------|-----------|
| 20-20-2-2    | 0.94 | 38 | 129 | 2.14 | 33 | 88 | **0.75** | 50 | 110 | 3.47 |
| 20-20-3-2    | 1.66 | 76 | 450 | 3.09 | 59 | 361 | **0.97** | 83 | 311 | 19.58 |
| 20-20-2-3    | 2.00 | 58 | 151 | 2.42 | 44 | 117 | **1.30** | 73 | 120 | 5.48 |
| 40-40-2-2    | 13.23 | 220 | 839 | 4.06 | 56 | 576 | **3.75** | 114 | 514 | 46.42 |
| 40-40-3-2    | 68.09 | 1,192 | 6,339 | 15.28 | 196 | 2155 | **11.16** | 369 | 2323 | 496.77 |
| 40-40-2-3    | 64.44 | 335 | 964 | 20.88 | 60 | 448 | **11.02** | 115 | 538 | 146.52 |
| 60-60-2-2    | 69.95 | 514 | 1,694 | 29.36 | 116 | 1172 | **9.67** | 124 | 1527 | 220.33 |
| 60-60-3-2    | 777.56 | 5,631 | 33,312 | 79.88 | 334 | 4006 | **39.33** | 549 | 11678 | 3630.78 |
| 60-60-2-3    | 640.11 | 755 | 2,290 | 211.22 | 121 | 1268 | **94.92** | 156 | 1462 | 1399.27 |
| 80-80-2-2    | 353.55 | 1,240 | 3,601 | 65.49 | 122 | 2119 | **25.75** | 175 | 3134 | 817.75 |
| 80-80-3-2    | 13655.10 | 15,236 | 93,558 | 147.78 | 345 | 10278 | **146.78** | 941 | 24219 | LIMIT |
| 80-80-2-3    | 5181.42 | 1,538 | 4,480 | 384.99 | 142 | 2421 | **228.08** | 207 | 3387 | 6989.31 |
| 100-100-2-2  | 636.63 | 1,573 | 5,741 | 57.97 | 89 | 2834 | **44.95** | 176 | 3656 | 2087.59 |
| 100-100-3-2  | 13418.00 | 22,628 | 155,384 | 233.02 | 323 | 7972 | **190.53** | 772 | 33046 | LIMIT |
| 100-100-2-3  | 5469.91 | 2,143 | 6,943 | 384.00 | 105 | 3124 | **273.86** | 194 | 4066 | LIMIT |

Average 2690.17 3,545 21,058 109.44 143 2596 **72.19** 273 6006 N/A

Figure 2  Improvement in Optimality Gap by Valid Inequalities
6.2. Strength of Valid Inequalities

We perform three implementations of Algorithm 1 to apply the submodular inequalities (8) only, the bulge inequalities (14) only, and both, denoted by SC, BI, and SCBI, respectively. Since there is no exact solution approach for (S-CFLP) in the existing literature, to benchmark these implementations, we solve each instance by an enumeration method, which finds optimal leader locations by evaluating the objective function value \( \min_{y \in Y} L(x, y) \) for all \( x \in X \).

In the first group of instances, the number of candidate facilities (i.e., \(|J|\)) ranges from 20 to 100 and the number of customers (i.e., \(|I|\)) is set to be equal to the number of candidate facilities. The values of \( p \) and \( r \) vary as 2 or 3. Each instance is indicated by its scale \(|I|-|J|-p-r\). We report the computational results in Table 3 and Figure 2. In Table 3, “Time(s)” report the CPU seconds it takes to solve each instance to optimum and we highlight the shortest solution time for each instance in bold. We report “LIMIT” in case if an instance is not solved to optimum within 4 hours. In addition, we report the total number of valid inequalities added until an optimal solution is found and the size of the final branch-and-bound tree in columns “#Cuts” and “#Nodes,” respectively. We record the gaps between the final optimal objective value and the objective value of the best integer solution found after performing the first, third and tenth lazy callbacks in CPLEX, denoted by “Gap1”, “Gap3”, and “Gap10”, respectively. Note that for each round of callback, at most one cut is added in SC and in BI, and at most two cuts are added in SCBI. In Figure 2, we report the average of the gaps (solid line), as well as the second smallest and second largest gaps (shaded area) among all instances (see the data of these gaps in Table 10 in Appendix C).

From Table 3 and Figure 2, we observe the following about the submodular and bulge inequalities.

1. Both inequalities strengthen the formulation significantly. For example, all three implementations outperform the benchmark method in all instances. In 10 rounds of cuts, all implementations are able to prove an optimality gap of below 3% in all instances (below 1% on average). In particular, the incorporation of both inequalities (i.e., the SCBI implementation) exhibits the best strength. For example, SCBI proves an average optimality gap of 0.65% and 0.20% across all instances in three and ten rounds of cuts, respectively (that is, by adding at most 6 and 20 cuts, respectively).

2. The submodular inequalities are less effective than the bulge inequalities. For example, BI solves each instance to optimum within 600 seconds. In contrast, SC spends 2690.17 seconds on average to solve an instance, which is roughly 25 times that of BI. In addition, SC involves more cuts and a larger branch-and-bound tree, which are roughly 25 and 8 times those of BI, respectively.
### Table 4  Computational results of instances with $|I| = |J| = 20$ and varying $p_r$, $r$-values

| Instance | SC | BI | SCBI |
|----------|----|----|------|
|          | #Cuts | Time(s) | Gap$_{10}$ | #Cuts | Time(s) | Gap$_{10}$ | #Cuts | Time(s) | Gap$_{10}$ | #Cuts |
| 20-20-2-2 | 38 | 0.97 | 0.00% | 33 | 0.56 | 0.00% | 50 |
| 20-20-4-2 | 140 | 2.70 | 0.00% | 95 | 1.16 | 0.00% | 117 |
| 20-20-6-2 | 350 | 6.61 | 1.49% | 90 | 1.75 | 0.23% | 201 |
| 20-20-8-2 | 194 | 3.58 | 2.58% | 150 | 1.39 | 3.04% | 170 |
| 20-20-10-2 | 132 | 2.67 | 0.00% | 121 | 1.44 | 0.00% | 182 |
| 20-20-2-4 | 70 | 3.09 | 0.00% | 45 | 1.34 | 0.00% | 74 |
| 20-20-4-4 | 299 | 9.14 | 4.23% | 175 | 3.88 | 4.20% | 310 |
| 20-20-6-4 | 504 | 13.27 | 2.36% | 353 | 6.05 | 3.81% | 578 |
| 20-20-8-4 | 711 | 17.11 | 0.00% | 781 | 9.75 | 0.00% | 953 |
| 20-20-10-4 | 881 | 17.45 | 1.93% | 1206 | 34.64 | 1.93% | 1552 |
| 20-20-2-6 | 75 | 2.73 | 2.08% | 46 | 1.75 | 0.00% | 82 |
| 20-20-4-6 | 395 | 11.36 | 5.64% | 292 | 8.52 | 6.49% | 497 |
| 20-20-6-6 | 981 | 24.50 | 6.46% | 1218 | 27.06 | 4.04% | 1441 |
| 20-20-8-6 | 1909 | 43.13 | 2.52% | 3453 | 57.06 | 2.52% | 3464 |
| 20-20-10-6 | 2885 | 50.69 | 0.93% | 5885 | 69.58 | 1.07% | 4399 |
| 20-20-2-8 | 81 | 2.64 | 1.68% | 58 | 1.58 | 1.68% | 123 |
| 20-20-4-8 | 519 | 12.70 | 9.36% | 620 | 8.30 | 9.36% | 821 |
| 20-20-6-8 | 1867 | 43.05 | 5.28% | 3308 | 33.94 | 5.40% | 3330 |
| 20-20-8-8 | 3067 | 68.17 | 2.54% | 9930 | 69.72 | 3.46% | 5579 |
| 20-20-10-8 | 2526 | 61.42 | 0.90% | 28118 | 81.05 | 0.00% | 4857 |
| 20-20-2-10 | 89 | 2.24 | 0.00% | 77 | 1.66 | 0.00% | 152 |
| 20-20-4-10 | 728 | 15.17 | 3.83% | 1065 | 11.14 | 0.18% | 1236 |
| 20-20-6-10 | 2133 | 45.11 | 7.26% | 6677 | 40.56 | 4.27% | 3962 |
| 20-20-8-10 | 3216 | 79.94 | 4.95% | 28190 | 92.58 | 4.95% | 6145 |
| 20-20-10-10 | 2964 | 74.02 | 2.00% | 1218 | 27.06 | 0.00% | 781 |
| Average | 1046 | 24.54 | 2.64% | 5026 | 26.17 | 2.27% | 1842 |

### Table 5  Computational results of instances with $|I| = |J| = 30$ and varying $p_r$, $r$-values

| Instance | SC | BI | SCBI |
|----------|----|----|------|
|          | #Cuts | Time(s) | Gap$_{10}$ | #Cuts | Time(s) | Gap$_{10}$ | #Cuts | Time(s) | Gap$_{10}$ |
| 30-30-3-3 | 567 | 34.11 | 0.99% | 141 | 6.06 | 3.10% | 214 |
| 30-30-6-3 | 5149 | 450.97 | 2.45% | 609 | 19.16 | 1.84% | 1021 |
| 30-30-9-3 | 11842 | 36.02 | 0.58% | 1361 | 34.19 | 1.62% | 2047 |
| 30-30-12-3 | 3723 | 2198.05 | 2.45% | 1245 | 20.20 | 2.45% | 1331 |
| 30-30-15-3 | 1130 | 88.42 | 2.07% | 1510 | 20.49 | 2.07% | 1409 |
| 30-30-3-6 | 825 | 143.83 | 0.17% | 245 | 34.53 | 0.17% | 493 |
| 30-30-6-6 | 12203 | 3259.80 | 1.45% | 2446 | 130.28 | 1.45% | 4425 |
| 30-30-9-6 | 14238 | LIMIT | N/A | 6131 | 400.83 | 0.69% | 9737 |
| 30-30-12-6 | 14453 | LIMIT | N/A | 18120 | 1981.98 | 0.25% | 7798 |
| 30-30-3-9 | 897 | 92.06 | 2.36% | 372 | 31.13 | 4.09% | 697 |
| 30-30-6-9 | 12108 | 2098.56 | 0.37% | 5487 | 304.11 | 0.37% | 8989 |
| 30-30-12-9 | 1919 | 55.72 | 1.09% | 493 | 20.95 | 3.88% | 898 |
| 30-30-15-9 | 14509 | 2199.31 | 1.76% | 14695 | 965.25 | 1.76% | 17479 |
| 30-30-3-15 | 968 | 39.92 | 4.01% | 715 | 21.39 | 4.01% | 1156 |
| 30-30-6-15 | 17089 | 2609.08 | 2.33% | 2330 | 131.13 | 2.10% | 3176 |

Average | 4417 | 888.40 | 1.60% | 140.74 | 1.85% | 2330 | 131.13 | 2.10% | 3176 |
3. SCBI performs even better than BI. For example, the average solution time by SCBI is about 34% less than that by BI. This suggests that the submodular and bulge inequalities complement each other quite well.

Next, we increase the values of \( p \) and \( r \) while fixing \(|I| = |J|\) at 20 and 30, respectively. Since search spaces of the leader’s and the follower’s problems increase exponentially with \( p \) and \( r \), this makes (S-CFLP) drastically more difficult to solve. We report the computational results in Tables 4–5. The instances that were not solved to optimum within 1 hour are marked as “LIMIT,” and if an implementation was not able to find any feasible solution in 10 rounds of cuts, we report “N/A” for \( \text{Gap}_{10} \). The average metrics in Tables 4–5 (reported in the bottom rows) are calculated among the instances that were solved to optimum by all three implementations.

From Tables 4–5, we observe that SCBI remains the most competitive implementation among the three alternatives. For example, the solution time of SCBI is the shortest in a majority of instances, and in the instances this is not the case, SCBI performs comparably with the best implementation. In contrast, the solution time of SC and BI increases quickly as \( p \) and \( r \) increase (see BI in Table 4 and SC in Table 5), reaching the time limit in several instances. This indicates that the submodular and bulge inequalities still complement each other well in these more challenging instances. For this reason, we adopt SCBI as the benchmark approach in all subsequent experiments.

### Table 6   Effectiveness of the approximate separation procedure

| Instance | SC | SC-AS | BI | BI-AS | SCBI | SCBI-AS |
|----------|----|-------|----|-------|------|---------|
| Time(s)  |    |       | Time(s) |       | Time(s) |       |
| Imp(%)   |    |       | Imp(%) |       | Imp(%) |       |
| 20-20-2-2| 0.94| 0.13  | 650.40  | 2.14  | 0.11  | 1864.22 |
|          |    |       |        |       | 0.75  |         |
|          |    |       |        |       | 0.13  | 500.00  |
| 20-20-3-2| 1.66| 0.20  | 715.76  | 3.09  | 0.23  | 1221.79 |
|          |    |       |        |       | 0.97  |         |
|          |    |       |        |       | 0.27  | 264.20  |
| 20-20-2-3| 2.00| 0.47  | 327.35  | 2.42  | 0.28  | 761.92  |
|          |    |       |        |       | 1.23  |         |
|          |    |       |        |       | 0.38  | 245.87  |
| 40-40-2-2| 13.23| 5.28  | 150.60  | 4.06  | 2.84  | 42.86  |
|          |    |       |        |       | 3.75  |         |
|          |    |       |        |       | 2.05  | 83.19   |
| 40-40-3-2| 68.09| 22.89 | 197.47  | 15.28 | 4.89  | 212.43  |
|          |    |       |        |       | 11.16 |         |
|          |    |       |        |       | 7.59  | 46.92   |
| 40-40-2-3| 64.44| 17.80 | 262.07  | 20.88 | 5.81  | 259.17  |
|          |    |       |        |       | 11.02 |         |
|          |    |       |        |       | 6.66  | 65.50   |
| 60-60-2-2| 69.95| 36.25 | 92.97   | 29.36 | 7.97  | 268.43  |
|          |    |       |        |       | 9.67  |         |
|          |    |       |        |       | 5.53  | 74.87   |
| 60-60-3-2| 777.56| 563.89| 37.90 | 79.88 | 25.06 | 218.70  |
|          |    |       |        |       | 39.33 |         |
|          |    |       |        |       | 29.44 | 33.60   |
| 60-60-2-3| 640.11| 148.75| 330.33 | 211.22| 40.27 | 424.57  |
|          |    |       |        |       | 94.92 |         |
|          |    |       |        |       | 33.03 | 187.37  |
| 80-80-2-2| 353.55| 244.41| 44.66 | 65.49 | 30.55 | 114.37  |
|          |    |       |        |       | 25.75 |         |
|          |    |       |        |       | 24.05 | 7.09    |
| 80-80-3-2| 13655.10| 6366.08| 114.50| 147.78| 57.36 | 157.64  |
|          |    |       |        |       | 146.78|         |
|          |    |       |        |       | 83.17 | 76.48   |
| 80-80-2-3| 5181.42| 2848.00| 81.93 | 384.99| 238.14| 61.66   |
|          |    |       |        |       | 228.08|         |
|          |    |       |        |       | 187.55| 21.61   |
| 100-100-2-2| 636.63| 356.59| 78.53 | 57.97 | 50.42 | 14.97  |
|          |    |       |        |       | 44.95 |         |
|          |    |       |        |       | 47.77 | -5.89   |
| 100-100-3-2| 13418.00| 14942.20| -10.20| 233.02| 102.11| 128.20  |
|          |    |       |        |       | 190.53|         |
|          |    |       |        |       | 175.80| 8.38    |
| 100-100-2-3| 5469.91| 4593.98| 19.07 | 384.00| 299.56| 28.19   |
|          |    |       |        |       | 278.36|         |
|          |    |       |        |       | 260.83| 5.00    |
| **Average** | **2690.17** | **2009.79** | **206.22** | **109.44** | **57.71** | **385.28** | **72.19** | **57.61** | **107.62** |
6.3. Effects of Approximate Separation

We evaluate the approximate separation method proposed in Section 4.3. That is, in line 6 of Algorithm 1, we always solve the approximate separation problem first and, only when this fails to find a cut, we invoke the “exact” separation to guarantee optimality of the algorithm. We denote these implementations as SC-AS, BI-AS, and SCBI-AS corresponding to SC, BI, and SCBI in Section 6.2. Table 6 reports the results, where \( \text{Imp}(\%) \) represents the relative improvement on solution time over the benchmark implementations. From this table, we observe that approximate separation speeds up Algorithm 1 significantly. For example, even for the most effective SCBI implementation, approximate separation provides a 2× speedup on average. We repeat this experiment on the more challenging instances with larger \( p \) and \( r \) values. The results are reported in Tables 11–12 in Appendix D and the observations are similar.

| Instance   | SCBI   | AA   | T-GAP | Obj-GAP |
|------------|--------|------|-------|---------|
| 20-20-2-2  | 0.75   | 0.5195 | 0.67  | 0.4768  |
| 20-20-3-2  | 0.97   | 0.2056 | 1.08  | 0.5917  |
| 20-20-3-3  | 1.30   | 0.4136 | 0.86  | 0.3707  |
| 40-40-2-2  | 3.75   | 0.5005 | 4.63  | 0.5000  |
| 40-40-3-2  | 11.16  | 0.6084 | 6.92  | 0.5878  |
| 40-40-3-3  | 11.02  | 0.3996 | 4.69  | 0.3993  |
| 60-60-2-2  | 9.67   | 0.5054 | 19.31 | 0.4889  |
| 60-60-3-2  | 39.33  | 0.0029 | 69.88 | 0.5978  |
| 60-60-3-3  | 94.92  | 0.4031 | 21.86 | 0.3901  |
| 80-80-2-2  | 25.75  | 0.5003 | 25.89 | 0.4941  |
| 80-80-3-2  | 146.78 | 0.5060 | 165.27| 0.5964  |
| 80-80-3-3  | 228.08 | 0.3954 | 28.53 | 0.3901  |
| 100-100-2-2| 44.95  | 0.4014 | 115.77| 0.5014  |
| 100-100-3-2| 190.53 | 0.0049 | 207.72| 0.6015  |
| 100-100-3-3| 273.86 | 0.3970 | 130.64| 0.3965  |
| Average   | 72.19  | 0.5055 | 12.89 | 0.4931  |

Table 8  Performance of the approximation algorithm on instances with \( |I| = |J| = 20 \) and varying \( p-, r \)-values

| Instance   | SCBI   | AA   | T-GAP | Obj-GAP |
|------------|--------|------|-------|---------|
| 20-20-2-2  | 0.56   | 0.5195 | 6.72  | 0.4768  |
| 20-20-3-2  | 1.16   | 0.6963 | 5.52  | 0.6584  |
| 20-20-3-3  | 1.75   | 0.7693 | 6.49  | 0.7585  |
| 20-20-4-2  | 1.39   | 0.8254 | 6.70  | 0.8160  |
| 20-20-4-3  | 1.34   | 0.3400 | 8.86  | 0.3392  |
| 20-20-4-4  | 3.88   | 0.5248 | 5.91  | 0.5096  |
| 20-20-4-6  | 6.05   | 0.6344 | 7.61  | 0.6149  |
| 20-20-5-2  | 1.75   | 0.7024 | 8.13  | 0.6860  |
| 20-20-5-3  | 3.88   | 0.5248 | 6.70  | 0.5402  |
| 20-20-5-4  | 6.05   | 0.6344 | 7.61  | 0.6149  |
| 20-20-5-6  | 7.52   | 0.4555 | 8.92  | 0.4079  |
| 20-20-5-8  | 10.62  | 0.5680 | 11.25 | 0.5680  |
| 20-20-6-2  | 1.75   | 0.2675 | 8.13  | 0.2675  |
| 20-20-6-3  | 5.52   | 0.4555 | 8.92  | 0.4079  |
| 20-20-6-4  | 7.52   | 0.4555 | 8.92  | 0.4079  |
| 20-20-6-5  | 10.62  | 0.5680 | 11.25 | 0.5680  |
| Average   | 26.17  | 0.5415 | 12.89 | 0.5224  |
6.4. Performance of the Approximation Algorithm

We evaluate the effectiveness of the MISOCPr approximation algorithm proposed in Section 5. To do this, we solve the same instances as in Section 6.2 by the SCBI implementation and by the MISOCPr formulation (15), respectively. We evaluate solution quality of the approximate solution by \( \text{Obj-GAP} := 1 - z^H / z^* \) and the relative saving on CPU seconds by \( \text{T-Gap} \) in Tables 7–8. From these tables, we observe that the proposed approximation algorithm is able to find near-optimal solutions in most instances. For example, the approximate algorithm produces a \( \text{Obj-GAP} \) below 5% in 78% of the instances (43/55). In addition, the algorithm achieves so within an average solution time roughly 45% of that of SCBI across all instances. Notably, the saving on solution time is particularly significant on challenging instances (e.g., 30-30-12-6 and 30-30-6-15). This demonstrates the applicability of the approximation algorithm in larger-sized instances.

6.5. Sensitivity Analysis

We analyze solution time and results of (S-CFLP) under various parameters settings. The results will provide insights on winning market share and how to conduct parameter selection in sequential CFLP.

![Graph showing solution time and optimal objective values given varying \( \beta \)](image)

**Figure 3** Solution time and optimal objective values given varying \( \beta \)

6.5.1. Varying \( \beta \) in Customer Choice Model  
Recall that parameter \( \beta \) measures the impact of traveling distance \( d_{ij} \) on the utility of a customer \( i \) patronizing facility \( j \), i.e., \( w_{ij} = \exp\{\alpha_j - \beta d_{ij}\} \). We examine how the optimal objective value and solution time of (S-CFLP) change
under various choices of $\beta$. In specific, we implement SCBI on the instance 100-100-2-2 with $\beta$ ranging from 0.01 to 1. We report the resulting optimal objective values and solution time in Figure 3 and the optimal locations of some representative instances in Figure 4.

From Figure 3, we observe that $\beta$ has large impact on the solution time. The solution time is small for most selections of $\beta$, but it becomes significantly higher when $\beta$ ranges between 0.2 and 0.3. Perhaps more interestingly, we observe that the optimal objective value has an increasing trend as $\beta$ increases, implying that the leader can widen its market share lead as customers become more unwilling to travel far for patronizing. In particular, there appears to be a “leap” in the leader’s share once $\beta$ exceeds 0.4. This indicates that once customers become sufficiently unwilling to travel far, the leader has a chance of developing a dominating market share even if the follower chooses her locations optimally in response. This observation suggests the importance of locating early and optimally in such markets.

Figure 4 depicts optimal leader and follower locations with $\beta = 0.05, 0.1, 0.3$, and 0.5. From this figure, we first observe that the optimal locations are clustered when $\beta$ is small (e.g., $\beta = 0.05$) and they spread out when $\beta$ increases. Indeed, a small $\beta$ implies lower spatial impedance effect, by which the facilities tend to be located at the center of the region to attract customers from all directions. As $\beta$ increases, customers tend to patronize local facilities only. As a result, it becomes optimal (for both leader and follower) to spread out the facilities in order to avoid self-competition and cover as many customers as possible. Second, we observe that the follower tends to locate its facilities near the leader’s (see Figures 4(a)–4(c)), demonstrating the economies of agglomeration. Nevertheless, this is not the case when $\beta$ becomes sufficiently large (e.g., $\beta = 0.5$). In that case, since customers only patronize locally, the follower harvests more customers by locating its facilities far away from the leader’s. This observation is particularly relevant when traveling for shopping becomes inconvenient (e.g., in challenging weather) or risky (e.g., during a pandemic).

### 6.5.2. Varying Size of Customers $|I|$  
In (S-CFLP), a set $I$ of discrete nodes are selected to represent a group of customers, and each node can represent a residential community or a business center located in the considered area. Intuitively, the larger $I$ is, the more representative the model would become, which may also be more challenging to solve. We examine the trade-off between model fidelity and computational burden by varying the size of customers $|I|$. In specific, we evaluate the solution time ($\text{Time(s)}$) and optimal objective value ($\text{Obj}$) of instances $|I|$-100-2-2 with $|I|$ ranging from 20 to 2000 and report the results in Table 9. From this table, we observe that the solution time remains under 5 minutes even if we increase $|I|$ to 2000. Relatively, the solution time increases by roughly 15 times while $|I|$ increases by 100 times. This demonstrates the scalability of Algorithm 1. In addition, we observe that the optimal objective value remains stable when $|I| \geq 100$. This demonstrates that setting $|I| = 100$, as in previous sections, produces sufficiently representative instances.
Figure 4  Leader’s and follower’s optimal location solutions given varying $\beta$. Dots, squares, and circles represent customer nodes, leader’s locations, and follower’s locations, respectively.

| $|I|$  | 20  | 40  | 60  | 80  | 100 | 200 | 400 | 800 | 1200 | 1600 | 2000 |
|------|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|
| Time(s) | 19.00 | 49.52 | 29.23 | 54.80 | 44.95 | 65.69 | 85.70 | 130.50 | 284.80 | 254.91 | 279.97 |
| Obj   | 0.5021 | 0.5004 | 0.5013 | 0.5003 | 0.5014 | 0.5011 | 0.5007 | 0.5009 | 0.5001 | 0.5002 | 0.5072 |
| #Cuts | 161  | 224  | 135  | 225  | 176  | 197  | 184  | 176  | 228  | 187  | 164  |

7. Conclusion and Future Research

This paper provides exact and approximation algorithms for solving sequential CFLP with a probabilistic customer choice model. We adopt integer programming and robust optimization methodologies to develop an exact, branch-and-cut algorithm. In addition, we derive an approximation algorithm with a constant guarantee on the ensuing market share. Extensive computational exper-
iments generate insights on winning market share and demonstrate the effectiveness of the two valid inequalities we derive. For future research, the (S-CFLP) model can be further strengthened by incorporating additional facility designing decisions, including the attractiveness level of facilities (such as sizes and capacities). Another interesting direction is to consider uncertainty and asymmetric information in decision making. For example, the follower’s strategy and budget for location may be unveiled to the leader.

References

Alekseeva E, Kochetov Y, Plyasunov A (2015) An exact method for the discrete \((r|p)\)-centroid problem. *Journal of Global Optimization* 63(3):445–460.

Aros-Vera F, Marianov V, Mitchell JE (2013) p-hub approach for the optimal park-and-ride facility location problem. *European Journal of Operational Research* 226(2):277–285.

Ben-Akiva ME, Lerman SR, Lerman SR (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, volume 9 (MIT press).

Benati S, Hansen P (2002) The maximum capture problem with random utilities: Problem formulation and algorithms. *European Journal of Operational Research* 143(3):518–530.

Dan T, Marcotte P (2019) Competitive facility location with selfish users and queues. *Operations Research* 67(2):479–497.

Díaz-Báñez JM, Heredia M, Pelegrín B, Pérez-Lantero P, Ventura I (2011) Finding all pure strategy Nash equilibria in a planar location game. *European Journal of Operational Research* 214(1):91–98.

Drezner T, Drezner Z (1998) Facility location in anticipation of future competition. *Location Science* 6(1-4):155–173.

Drezner T, Drezner Z, Kalczynski P (2015) A leader-follower model for discrete competitive facility location. *Computers and Operations Research* 64:51–59.

Duran MA, Grossmann IE (1986) An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming* 36(3):307–339.

Edmonds J (1970) Submodular functions, matroids, and certain polyhedra. *Combinatorial Structures and Their Applications* 69–87.

Eiselt HA, Laporte G (1997) Sequential location problems. *European Journal of Operational Research* 96(2):217–231.

Fischer K (2002) Sequential discrete p-facility models for competitive location planning. *Annals of Operations Research* 111(1-4):253–270.

Freire AS, Moreno E, Yushimito WF (2016) A branch-and-bound algorithm for the maximum capture problem with random utilities. *European Journal of Operational Research* 252(1):204–212.
Gentile J, Alves Pessoa A, Poss M, Costa Roboredo M (2018) Integer programming formulations for three sequential discrete competitive location problems with foresight. *European Journal of Operational Research* 265(3):872–881.

Godinho P, Dias J (2010) A two-player competitive discrete location model with simultaneous decisions. *European Journal of Operational Research* 207(3):1419–1432.

Godinho P, Dias J (2013) Two-player simultaneous location game: Preferential rights and overbidding. *European Journal of Operational Research* 229(3):663–672.

Haase K (2009) Discrete location planning. Technical report, Institute of Transport and Logistics Studies Working Paper ITLS-WP-09-07.

Haase K, Müller S (2014) A comparison of linear reformulations for multinomial logit choice probabilities in facility location models. *European Journal of Operational Research* 232(3):689–691.

Hakimi SL (1983) On locating new facilities in a competitive environment. *European Journal of Operational Research* 12(1):29–35.

Henrici P (1961) Two remarks on the Kantorovich inequality. *The American Mathematical Monthly* 68(9):904–906.

Huff DL (1964) Defining and estimating a trading area. *The Journal of Marketing* 28(3):34–38.

Kantorovich L (1948) Functional analysis and applied mathematics (in Russian). *Uspehi Mat. Nauk.* 3(28):89–185.

Kleinert T, Labbé M, Ljubić I, Schmidt M (2021) A survey on mixed-integer programming techniques in bilevel optimization. [http://www.optimization-online.org/DB_FILE/2021/01/8187.pdf](http://www.optimization-online.org/DB_FILE/2021/01/8187.pdf).

Krause A, McMahan HB, Guestrin C, Gupta A (2008) Robust submodular observation selection. *Journal of Machine Learning Research* 9:2761–2801.

Kress D, Pesch E (2012) Sequential competitive location on networks. *European Journal of Operational Research* 217(3):483–499.

Küçükaydın H, Aras N, Kuban Altınel I (2011) Competitive facility location problem with attractiveness adjustment of the follower: A bilevel programming model and its solution. *European Journal of Operational Research* 208:206–220.

Küçükaydın H, Aras N, Kuban Altınel I (2012) A leader-follower game in competitive facility location. *Computers and Operations Research* 39(2):437–448.

Ljubić I, Moreno E (2018) Outer approximation and submodular cuts for maximum capture facility location problems with random utilities. *European Journal of Operational Research* 266(1):46–56.

Luce RD (1959) *Individual Choice Behavior: A Theoretical Analysis* (Wiley).

Mai T, Lodi A (2020) A multicut outer-approximation approach for competitive facility location under random utilities. *European Journal of Operational Research* 284(3):874–881.
McFadden D (1973) Conditional logit analysis of qualitative choice behavior. Zarembka P, ed., *Frontiers in Econometrics (Economic Theory and Mathematical Economics)*, 105–142 (Academic Press).

Nemhauser GL, Wolsey LA (1981) Maximizing submodular set functions: formulations and analysis of algorithms. *Studies on Graphs and Discrete Programming* 11:279–301.

O’Kelly ME (1999) Trade-area models and choice-based samples: Methods. *Environment and Planning A* 31(4):613–627.

Plastria F (2001) Static competitive facility location: An overview of optimisation approaches. *European Journal of Operational Research* 129(3):461–470.

Plastria F, Vanhaverbeke L (2008) Discrete models for competitive location with foresight. *Computers and Operations Research* 35(3):683–700.

ReVelle C (1986) The maximum capture or “sphere of influence” location problem: Hotelling revisited on a network. *Journal of Regional Science* 26(2):343–358.

ReVelle C, Serra F (1995) Competitive location in discrete space. Drezner Z, ed., *Facility Location: A Survey of Applications and Methods* (Springer).

Roboredo MC, Pessoa AA (2013) A branch-and-cut algorithm for the discrete \((r | p)\)-centroid problem. *European Journal of Operational Research* 224(1):101–109.

Sáiz ME, Hendrix EM, Fernández J, Pelegrín B (2009) On a branch-and-bound approach for a Huff-like Stackelberg location problem. *OR Spectrum* 31(3):679–705.

Serra D, ReVelle C (1994) Market capture by two competitors: The preemptive location problem. *Journal of Regional Science* 34(4):549–561.

Slater D (1975) Underdevelopment and spatial inequality: Approaches to the problems of regional planning in the third world. *Progress in Planning* 4:97–167.

Topkis DM (1978) Minimizing a submodular function on a lattice. *Operations Research* 26(2):305–321, ISSN 0030364X, 15265463, URL http://www.jstor.org/stable/169636.

von Stackelberg H (1934) *Marktform und Gleichgewicht* (Springer).

Zhang Y, Berman O, Verter V (2012) The impact of client choice on preventive healthcare facility network design. *OR Spectrum* 34(2):349–370.

A. Proofs

A.1. Proof of Theorem 1

**Proof:** By Theorem 2, which we shall prove later, \((S\text{-CFLP})\) is equivalent to \(\max_{x \in X} \min_{y \in Y} L(x, y)\), where \(X = \{x \in \{0, 1\}^{|J|} : e^\top x \leq p\}\), \(Y = \{y \in \{0, 1\}^{|J|} : e^\top y \leq r\}\), and \(L(x, y)\) is defined in (4). In addition, for any fixed \(y \in Y\), \(L(x, y)\) is submodular in the index set \(X\) of variables \(x\) (defined as \(X := \{j \in J : x_j = 1\}\)) by Proposition 1 presented later. Hence, \((S\text{-CFLP})\) is equivalent to a
robust submodular maximization model as defined in formulation (2) of Krause et al. (2008). The conclusion follows from Theorem 3 of Krause et al. (2008). □

A.2. Proof of Proposition 1

Proof: For all \( i \in I, X, Y \subseteq J, \) and \( k \in J \setminus X \), it follows from (7) that

\[
\begin{align*}
  f_{i,Y}(X \cup \{k\}) - f_{i,Y}(X) &= \frac{U_i^L + \sum_{j \in X \cup \{k\}} w_{ij}}{U_i^L + U_i^F + \sum_{j \in X \cup Y \cup \{k\}} w_{ij}} - \frac{U_i^L + \sum_{j \in X} w_{ij}}{U_i^L + U_i^F + \sum_{j \in X \cup Y} w_{ij}} \\
  &= \begin{cases} 
    \frac{U_i^L + U_i^F + \sum_{j \in X \cup Y \setminus X} w_{ij}}{U_i^L + U_i^F + \sum_{j \in X \cup Y} w_{ij}} & \text{if } k \in Y \\
    \frac{U_i^L + U_i^F + \sum_{j \in X \cup Y \setminus X} w_{ij}}{(U_i^L + U_i^F + \sum_{j \in X \cup Y} w_{ij})} & \text{if } k \notin Y.
  \end{cases}
\end{align*}
\]

(16)

It follows that \( f_{i,Y}(X \cup \{k\}) - f_{i,Y}(X) \) is non-increasing in \( X \), i.e., \( f_{i,Y}(X \cup \{k\}) - f_{i,Y}(X) \geq f_{i,Y}(X' \cup \{k\}) - f_{i,Y}(X') \) for all \( X \subseteq X' \subseteq J \setminus \{k\} \). Therefore, \( f_{i,Y} \) is submodular and so is \( L_Y \) because \( L_Y \) is a linear combination of \( f_{i,Y} \). This completes the proof. □

A.3. Proof of Proposition 3

Proof: For notation brevity, we omit the subscript \( Y \) in this proof. Pick any subsets \( S \subseteq R \subseteq J \) and any element \( j \in J \setminus R \). By definition, we have

\[
H(S \cup \{j\}) = L(S \cup \{j\}) - \sum_{k \in S} \rho(J \setminus \{k\}; k)(1 - \hat{x}_k) - \rho(J \setminus \{j\}; j)(1 - \hat{x}_j) + \sum_{k \in J \setminus S} \rho(S; k)\hat{x}_k - \rho(S; j)\hat{x}_j.
\]

Then,

\[
H(S \cup \{j\}) - H(S) = L(S \cup \{j\}) - L(S) - \rho(J \setminus \{j\}; j)(1 - \hat{x}_j) - \rho(S; j)\hat{x}_j
\]

\[
= (\rho(S; j) - \rho(J \setminus \{j\}; j))(1 - \hat{x}_j).
\]

It follows that

\[
[H(S \cup \{j\}) - H(S)] - [H(R \cup \{j\}) - H(R)] = (\rho(S; j) - \rho(R; j))(1 - \hat{x}_j) \geq 0,
\]

where the inequality follows from Proposition 1. Therefore, \( H(S \cup \{j\}) - H(S) \) is non-increasing in \( S \) and thus, \( H \) is submodular by definition. This completes the proof. □

A.4. Proof of Proposition 5

Proof: By definition, \( \theta \leq \hat{L}(x, y) \) holds valid if and only if there exist \( \{\theta_i\}_{i \in I} \) such that \( \theta \leq \sum_{i \in I} h_i \theta_i \) and

\[
\theta_i \leq \frac{U_i^L + \sum_{j \in J} w_{ij} \left[-y_jx_j^2 + (1 + y_j)x_j\right]}{U_i^L + U_i^F + \sum_{j \in J} w_{ij} \left[(1 - y_j)x_j + y_j\right]}, \quad \forall i \in I.
\]

(17)
To show that inequality (17) is second-order conic representable, we note that

\[
\sum_{j \in J} w_{ij} [-y_j x_j^2 + (1 + y_j) x_j] = \sum_{j \in J_0} w_{ij} x_j + \sum_{j \in J_1} w_{ij} (-x_j^2 + 2x_j) \quad \text{and} \quad \sum_{j \in J} w_{ij} [(1 - y_j) x_j + y_j] = \sum_{j \in J_0} w_{ij} x_j + \sum_{j \in J_1} w_{ij} \cdot 1 > 0.
\]

We finish the proof by rewriting inequality (17) as follows.

\[
(17) \iff \theta_i \left( U_i^L + U_i^P + \sum_{j \in J_0} w_{ij} x_j + \sum_{j \in J_1} w_{ij} \right) \leq U_i^L + U_i^P + \sum_{j \in J_0} w_{ij} x_j + \sum_{j \in J_1} w_{ij} = U_i^P
\]

\[
\iff \sum_{j \in J_1} (1 - x_j)^2 + \left( \frac{\sqrt{U_i^P}}{1} \right)^2 \leq \left( U_i^L + U_i^P + \sum_{j \in J_0} w_{ij} x_j + \sum_{j \in J_1} w_{ij} \right) (1 - \theta_i)
\]

\[
\iff \sum_{j \in J_1} (1 - x_j)^2 + \left( \frac{\sqrt{U_i^P}}{1} \right)^2 + \left( \sum_{j \in J} w_{ij} (1 - y_j) x_j + s_i(y) + \theta_i - 1 \right)^2
\]

\[
\leq \left( \sum_{j \in J} w_{ij} (1 - y_j) x_j + s_i(y) - \theta_i + 1 \right) \iff (13). \quad \square
\]

A.5. Proof of Lemma 1

Proof: First, we recall that (S-CFLP) is equivalent to \( \min_{x \in X, y \in Y} \min L(x, y) \). Suppose that \( \bar{x} \) is an optimal solution to (S-CFLP) and \( e^T \bar{x} < p \). Then, there exists a \( k \in J \) such that \( \bar{x}_k = 0 \). We construct a new solution \( \bar{x}' \) such that \( \bar{x}'_k = 1 \) and \( \bar{x}_j = \bar{x}'_j \) for all \( j \neq k \). Then, \( \bar{x}' \in X \) because \( e^T \bar{x}' \leq p \). In addition, for any \( y \in Y \), we notice that \( L(\bar{x}', y) \geq L(\bar{x}, y) \) by discussing the following two cases.

1. If \( y_k = 1 \), then \( (\bar{x}_j \vee y_j) = (\bar{x}'_j \vee y_j) \) for all \( j \in J \). It follows that

\[
L(\bar{x}', y) = \sum_{i \in I} h_i \frac{U_i^L + \sum_{j \in J \setminus \{k\}} w_{ij} \bar{x}'_j + w_{ik}}{U_i^L + U_i^P + \sum_{j \in J} w_{ij} (\bar{x}'_j \vee y_j)}
\]

\[
= \sum_{i \in I} h_i \frac{U_i^L + \sum_{j \in J \setminus \{k\}} w_{ij} \bar{x}_j + w_{ik}}{U_i^L + U_i^P + \sum_{j \in J} w_{ij} (\bar{x}_j \vee y_j)} = L(\bar{x}, y).
\]

2. If \( y_k = 0 \), then \( (\bar{x}_k \vee y_k) = 0, (\bar{x}'_k \vee y_k) = 1, \) and \( (\bar{x}_k \vee y_k) - \bar{x}_k = 0 = (\bar{x}'_k \vee y_k) - \bar{x}'_k \). It follows that

\[
L(\bar{x}', y) = 1 - \sum_{i \in I} h_i \frac{U_i^P + \sum_{j \in J} w_{ij} [(\bar{x}'_j \vee y_j) - \bar{x}'_j]}{U_i^L + U_i^P + \sum_{j \in J \setminus \{k\}} w_{ij} (\bar{x}'_j \vee y_j) + w_{ik} (\bar{x}'_k \vee y_k)}
\]

\[
= 1 - \sum_{i \in I} h_i \frac{U_i^P + \sum_{j \in J} w_{ij} [(\bar{x}_j \vee y_j) - \bar{x}_j]}{U_i^L + U_i^P + \sum_{j \in J \setminus \{k\}} w_{ij} (\bar{x}_j \vee y_j) + w_{ik}}
\]

\[
\geq 1 - \sum_{i \in I} h_i \frac{U_i^P + \sum_{j \in J} w_{ij} [(\bar{x}_j \vee y_j) - \bar{x}_j]}{U_i^L + U_i^P + \sum_{j \in J \setminus \{k\}} w_{ij} (\bar{x}_j \vee y_j)} = L(\bar{x}, y).
\]
Therefore, \( \min_{y \in Y} L(\tilde{x}, y) \geq \min_{y \in Y} L(\bar{x}, y) \) and so \( \bar{x}' \) is also optimal. Repeating this procedure yields an optimal solution \( x^* \) to (S-CFLP) such that \( e^\top x^* = p \).

Second, for any fixed \( \tilde{x} \in \mathcal{X} \), suppose that \( \bar{y} \) is an optimal solution to the separation problem (6) such that \( e^\top \bar{y} < r \). Then, we can construct a \( y^* \in Y \) with \( e^\top y^* = r \) by flipping sufficiently many entries of \( \bar{y} \) from zero to one. Since \((\hat{x}_j \lor \bar{y}_j) \leq (\hat{x}_j \lor y^*_j)\) for all \( j \in J \), we have \( L(\hat{x}, y^*) \leq L(\hat{x}, \bar{y}) \). Therefore, \( y^* \) is also optimal to (6). \( \square \)

A.6. Proof of Proposition 6

Proof: First, we rewrite \( L(\hat{x}, y) \) as

\[
L(\hat{x}, y) = \sum_{i \in I} h_i \left( \frac{U_i^L + \sum_{j \in J} w_{ij} \hat{x}_j}{U_i^L + U_i^P + \sum_{j \in J} w_{ij}(\hat{x}_j + y_j - \hat{x}_j \cdot y_j)} \right)
\]

where \( w_i(\hat{x}) := U_i^P + \sum_{j \in J} w_{ij}(1 - \hat{x}_j)y_j \) for all \( i \in I \). Without loss of optimality, we can assume that \( w_i^L(\hat{x}) \leq w_i(\hat{x}) \leq w_i^U(\hat{x}) \). Since \( a_i(\hat{x}) > 0 \) and \( w_i^L(\hat{x}) \geq 0 \), the function \( \ell(w) := \frac{a_i(\hat{x})}{a_i(\hat{x}) + w} \) is convex in the interval \([w_i^L(\hat{x}), w_i^U(\hat{x})]\). As a result,

\[
L(\hat{x}, y) \equiv \ell(w_i(\hat{x})) \leq \sum_{i \in I} h_i \left[ \ell(w_i^L(\hat{x})) + \frac{\ell(w_i^U(\hat{x})) - \ell(w_i^L(\hat{x}))}{w_i^L(\hat{x}) - w_i^U(\hat{x})(w_i(\hat{x}) - w_i^L(\hat{x}))} \right] w_i(\hat{x})
\]

\[
= \sum_{i \in I} h_i \left[ \frac{a_i(\hat{x})(a_i(\hat{x}) + w_i^L(\hat{x}) + w_i^U(\hat{x}))}{a_i(\hat{x}) + w_i^L(\hat{x})(a_i(\hat{x}) + w_i^U(\hat{x}))} - \frac{a_i(\hat{x})}{(a_i(\hat{x}) + w_i^L(\hat{x})(a_i(\hat{x}) + w_i^U(\hat{x}))} \right] w_i(\hat{x})
\]

\[
= \alpha(\hat{x}) - \beta(\hat{x})^\top y.
\]

Second, since \( \alpha(\hat{x}) - \beta(\hat{x})^\top y \) is affine in \( y \), a greedy algorithm solves the problem \( \min_{y \in Y} \{ \alpha(\hat{x}) - \beta(\hat{x})^\top y \} \) with an optimal solution \( \hat{y} \) as described in the claim of this proposition. \( \square \)

A.7. Proof of Theorem 3

Proof: First, by Lemma 1, (S-CFLP) can be represented as \( \max_{x \in X} \min_{y \in Y} L(x, y) \). For all \( x \in X, y \in Y \), and \( i \in I \), we define \( a_i(x) := U_i^L + \sum_{j \in J} w_{ij}x_j \) and \( w_i(x, y) := U_i^P + \sum_{j \in J} w_{ij}(1 - x_j)y_j \). Then, \( L(x, y) = \sum_{i \in I} h_i \left( \frac{a_i(x)}{a_i(x) + w_i(x, y)} \right) \). We bound \( L(x, y) \) from below by the harmonic-geometric-arithmetic mean inequality:

\[
L(x, y) \geq \frac{1}{\sum_{i \in I} h_i \left( \frac{a_i(x) + w_i(x, y)}{a_i(x)} \right)}, \quad \forall x \in X, \forall y \in Y.
\]

(18)
It follows that
\[
\begin{align*}
\max & \min_{x \in \mathcal{X}} \min_{y \in \mathcal{Y}} L(x, y) \\
= & \max_{x \in \mathcal{X}} \min_{y \in \mathcal{Y}} \sum_{i \in I} h_i \left( \frac{a_i(x) + w_i(x, y)}{a_i(x)} \right) \\
= & \frac{1}{1 + \min \max_{x \in \mathcal{X}, y \in \mathcal{Y}} \sum_{i \in I} \left( \frac{h_i w_i(x, y)}{a_i(x)} \right)}.
\end{align*}
\]
That is, solving problem
\[
\min \max_{x \in \mathcal{X}, y \in \mathcal{Y}} \sum_{i \in I} \left( \frac{h_i w_i(x, y)}{a_i(x)} \right)
\]
produces a feasible solution and a lower bound of (S-CFLP).

Second, we recast formulation (20) as the MISOP (15). To do this, we derive
\[
\begin{align*}
\min & \max_{x \in \mathcal{X}, \mu \geq 0} \sum_{i \in I} \left( \frac{h_i w_i(x, y)}{a_i(x)} \right) \\
= & \min_{x \in \mathcal{X}, \mu \geq 0} \left\{ \sum_{i \in I} \left( \frac{h_i w_i(x, y)}{a_i(x)} \right) + \sum_{j \in J} \left( \sum_{i \in I} h_i w_i \left( 1 - x_j \right) \right) y_j \right\} \\
= & \min_{x \in \mathcal{X}, \mu \geq 0, s \geq 0, t \geq 0} \sum_{i \in I} h_i w_i s_i + r \lambda + e^\top \mu \\
\quad \text{s.t.} & \lambda + \mu_j \geq \sum_{i \in I} \left( \frac{h_i w_i \left( 1 - x_j \right) \right)}{U_i^l + \sum_{k \in J} w_{ik} x_k}, \forall j \in J, \\
\quad & s_i \geq \frac{1}{U_i^l + \sum_{k \in J} w_{ik} x_k}, \forall i \in I, \\
\quad & t_{ij} \geq \frac{1 - x_j}{U_i^l + \sum_{k \in J} w_{ik} x_k}, \forall i \in I, \forall j \in J.
\end{align*}
\]
where equality (21b) is due to the strong duality of linear programming. Indeed, since the inner maximization problem in (21a) has an objective function linear in $y$, the feasible region $\mathcal{Y}$ of $y$ can be replaced by its convex hull $\text{conv}(\mathcal{Y}) = \{ y \in [0, 1]^{\mathcal{J}} : e^\top y = r \}$, because $e^\top y = r$ produces a totally unimodular constraint matrix. Then, this inner maximization problem is equivalent to a linear program with a feasible region $y \in \{ y \in \mathbb{R}_+^{\mathcal{J}} : e^\top y = r, y \leq e \}$. Taking the dual of this linear program yields formulation (21c)–(21f), where dual variables $\lambda$ and $\mu$ are associated with primal constraints $e^\top y = r$ and $y \leq e$, respectively. We represent constraints (21e) as
\[
(21e) \iff \left( s_i + U_i^l + \sum_{k \in J} w_{ik} x_k \right)^2 - \left( U_i^l + \sum_{k \in J} w_{ik} x_k - s_i \right)^2 \geq 4
\]
\[
\iff \left\| U_i^l + \sum_{k \in J} w_{ik} x_k - s_i \right\| \leq s_i + U_i^l + \sum_{k \in J} w_{ik} x_k \iff (15c).
\]
Likewise, we recast constraints (21f) as

\[(21f) \iff t_{ij} \geq \frac{(1-x_j)^2}{U_i^L + \sum_{k \in J} w_{ik} x_k}\]

\[\iff \|\left[U_i^L + \sum_{k \in J} w_{ik} x_k - t_{ij}\right]\| \leq t_{ij} + U_i^L + \sum_{k \in J} w_{ik} x_k \iff (15d),\]

where the first equivalence is because \(x_j\) is binary-valued.

Third, let \(\ell(x,y) := \left[\sum_{i \in I} h_i \left(\frac{a_i(x) + w_i(x,y)}{a_i(x)}\right)\right]^{-1}\). Then, \(x^H \in \arg\max_{x \in X} \left\{\min_{y \in Y} \ell(x,y)\right\}\) by the equivalence between (20) and (15). In addition, \(L(x,y) \geq \ell(x,y)\) by inequality (18). We bound \(L(x,y)\) from above by deriving

\[L(x,y) = \sum_{i \in I} h_i \left(\frac{a_i(x)}{a_i(x) + w_i(x,y)}\right) \leq \left(\frac{\gamma_M(x,y) + \gamma_l(x,y)}{4\gamma_M(x,y)\gamma_l(x,y)}\right) \left[\sum_{i \in I} h_i \left(\frac{a_i(x) + w_i(x,y)}{a_i(x)}\right)\right]^{-1}\]

\[\leq \left(\frac{\gamma_M + \gamma_l}{4\gamma_M\gamma_l}\right) \ell(x,y),\]

where \(\gamma_M(x,y) := \max_{i \in I} \left\{\frac{a_i(x)}{a_i(x) + w_i(x,y)}\right\}\) and \(\gamma_l(x,y) := \min_{i \in I} \left\{\frac{a_i(x) + w_i(x,y)}{a_i(x)}\right\}\). The first inequality above follows from the Kantorovich inequality (see Kantorovich 1948, Henrici 1961), and the second inequality is because

\[\frac{(\gamma_M(x,y) + \gamma_l(x,y))^2}{4\gamma_M(x,y)\gamma_l(x,y)} = \frac{1}{4} \left(\frac{\gamma_M(x,y)}{\gamma_l(x,y)} + \frac{\gamma_l(x,y)}{\gamma_M(x,y)} + 2\right) \leq \frac{1}{4} \left(\frac{\gamma_M}{\gamma_l} + \frac{\gamma_l}{\gamma_M} + 2\right) = \left(\frac{\gamma_M + \gamma_l}{4\gamma_M\gamma_l}\right)^2,\]

where the inequality is because \(1 \leq \gamma_M(x,y) \leq \frac{2\gamma_M}{\gamma_l}\) and function \(f(z) := z + \frac{1}{z} + 2\) is increasing on the interval \([1, \infty)\).

Now pick any \(x^* \in \arg\max_{x \in X} \left\{\min_{y \in Y} L(x,y)\right\}\). Then, \(z^H = \min_{y \in Y} L(x^H, y) \leq \min_{y \in Y} L(x^*, y) = z^*\) by definition of \(x^*\). In addition,

\[z^H = \min_{y \in Y} L(x^H, y) \geq \min_{y \in Y} \ell(x^H, y) \geq \min_{y \in Y} \left\{\frac{4\gamma_M\gamma_l}{(\gamma_M + \gamma_l)^2} L(x^*, y)\right\} = \frac{4\gamma_M\gamma_l}{(\gamma_M + \gamma_l)^2} z^*,\]

where the first and the third inequalities are because \(\ell(x,y) \leq L(x,y) \leq \frac{(\gamma_M + \gamma_l)^2}{4\gamma_M\gamma_l} \ell(x,y)\), and the second inequality is by definition of \(x^H\). This finishes the proof. \(\square\)

\section*{B. Valid Inequalities for MISOCOCP (15)}

We derive valid linear inequalities with regard to the second-order conic constraints (15c)–(15d) (more specifically, their representation (21e)–(21f)) to further improve the efficacy of solving MISOCOCP (15). These valid inequalities exploit the convexity of the right-hand side of inequality (21e),
and although that of \((21f)\) is non-convex, we recover the convexity by resorting to perspective function.

**Proposition 8.** Given any \(\hat{x} \in \mathcal{X}\), the following linear inequalities are valid for MISOCP \((15)\):

\[
\begin{align*}
    s_i & \geq \frac{U^L_i + \sum_{k \in J} w_{ik} (2\hat{x}_k - x_k)}{(U^L_i + \sum_{k \in J} w_{ik}\hat{x}_k)^2}, \quad \forall i \in I, \\
    t_{ij} & \geq \frac{(1 - \hat{x}_j) \left( (U^L_i + 2\sum_{k \neq j} w_{ik}\hat{x}_k)(1 - x_j) - \sum_{k \neq j} w_{ik} x_k \right)}{(U^L_i + \sum_{k \neq j} w_{ik}\hat{x}_k)^2}, \quad \forall i \in I, \forall j \in J.
\end{align*}
\]  

\((23a)\) \((23b)\)

**Proof of Proposition 8:** From the proof of Theorem 3, we observe that the second-order conic constraints \((15c)\) and \((15d)\) are equivalent to inequalities \((21e)\) and \((21f)\), respectively.

First, we notice that the right-hand side of inequality \((21e)\) is convex in variables \(x\) because it is the composite of function \(c(z) := 1/z\) on \(z > 0\) and an affine function \(z(x) := U^L_i + \sum_{k \in J} w_{ik} x_k\). Then, the supporting hyperplane of this convex function yields inequalities \((23a)\).

Second, we notice that the right-hand side of inequality \((21f)\) is not convex in variables \(x\). Nevertheless, for each \(j \in J\), this right-hand side is convex in variables \((x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_{|J|})\) when fixing \(x_j = 0\). In contrast, if we fix \(x_j = 1\) then this right-hand side becomes zero. This motivates us to replace the right-hand side of \((21f)\) with the perspective function of its restriction generated by fixing \(x_j = 0\). More specifically, we consider inequalities

\[
    t_{ij} \geq t_{ij}(x) := \frac{(1 - x_j)^2}{U^L_i(1 - x_j) + \sum_{k \neq j} w_{ik} x_k}, \quad \forall i \in I, \forall j \in J.
\]

We notice that this inequality is equivalent to \((21f)\) whenever \(x \in \mathcal{X} \subseteq \{0, 1\}^{|J|}\), but its right-hand side \(t_{ij}(x)\) is now convex in variables \(x\) because it is the perspective function of \(1/(U^L_i + \sum_{k \neq j} w_{ik} x_k)\), which is convex in \(x\). Then, the supporting hyperplane of \(t_{ij}(x)\) yields inequalities \((23b)\). This finishes the proof. \(\square\)

In computation, we incorporate inequalities \((23a)\)–\((23b)\) via the lazy callback.

**C. Improvement in Optimality Gaps by Valid Inequalities**

**D. Additional Results on Approximate Separation**
Table 10  Improvement in Optimality Gaps after Adding One, Three, and Ten Rounds of Valid Inequalities

| Instance   | SC  | BI  | SCBI |
|------------|-----|-----|------|
|            | Gap1 | Gap3 | Gap10 | Gap1 | Gap3 | Gap10 | Gap1 | Gap3 | Gap10 |
| 20-20-2-2  | 5.84% | 5.84% | 0.00% | 36.13% | 0.00% | 0.00% | 5.84% | 0.00% | 0.00% |
| 20-20-3-2  | 18.45% | 2.96% | 2.96% | 18.71% | 0.86% | 0.00% | 18.45% | 2.96% | 0.00% |
| 40-40-2-2  | 0.00% | 0.00% | 0.00% | 19.91% | 14.78% | 0.00% | 0.00% | 0.00% | 0.00% |
| 40-40-3-2  | 17.82% | 6.75% | 2.19% | 21.84% | 2.19% | 0.47% | 17.82% | 0.47% | 0.47% |
| 60-60-2-2  | 1.20% | 0.00% | 0.00% | 24.48% | 9.18% | 0.93% | 1.20% | 0.93% | 0.00% |
| 60-60-3-2  | 100.00% | 1.13% | 0.37% | 11.26% | 0.15% | 0.00% | 11.26% | 0.00% | 0.00% |
| 80-80-2-2  | 0.10% | 0.10% | 0.10% | 5.39% | 2.96% | 2.41% | 0.10% | 2.41% | 0.00% |
| 80-80-3-2  | 100.00% | 100.00% | 0.00% | 25.16% | 7.58% | 0.00% | 25.16% | 7.58% | 0.00% |
| 100-100-2-2 | 0.00% | 0.00% | 0.00% | 9.94% | 2.18% | 0.15% | 0.00% | 2.18% | 0.15% |
| 100-100-3-2 | 6.02% | 6.02% | 0.15% | 5.53% | 1.50% | 0.37% | 6.02% | 1.50% | 0.37% |
| Average    | 23.63% | 15.16% | 0.59% | 17.12% | 3.73% | 0.56% | 5.07% | 0.65% | 0.20% |

Table 11  Effectiveness of approximation separation on instances with |I| = |J| = 20 and varying p-, r-values

| Instance   | SC    | SC-AS | BI    | BI-AS | SCBI   | SCBI-AS |
|------------|-------|-------|-------|-------|--------|---------|
|            | Time(s) | Time(s) | Imp(%) | Time(s) | Time(s) | Imp(%) | Time(s) | Time(s) | Imp(%) |
| 20-20-2-2  | 0.97   | 0.16   | 521.15 | 0.88   | 0.13   | 600.00 | 0.56   | 0.13   | 349.60 |
| 20-20-4-2  | 2.70   | 0.89   | 203.37 | 2.03   | 0.28   | 622.78 | 1.16   | 0.75   | 54.13  |
| 20-20-6-2  | 6.61   | 3.63   | 82.34  | 1.69   | 0.67   | 151.19 | 1.75   | 1.25   | 40.00  |
| 20-20-8-2  | 3.58   | 2.30   | 55.77  | 2.45   | 0.61   | 302.13 | 1.39   | 0.81   | 71.18  |
| 20-10-2-2  | 2.67   | 1.36   | 96.62  | 1.91   | 0.80   | 139.15 | 1.44   | 0.81   | 76.88  |
| 20-20-4-2  | 3.09   | 0.99   | 214.11 | 1.67   | 0.44   | 281.74 | 1.34   | 0.50   | 168.80 |
| 20-20-6-2  | 9.14   | 3.84   | 137.77 | 4.19   | 0.56   | 643.87 | 3.88   | 1.48   | 161.12 |
| 20-20-10-2 | 14.27  | 5.08   | 161.24 | 6.83   | 1.50   | 355.20 | 6.05   | 3.02   | 100.50 |
| 20-20-2-4  | 17.11  | 8.13   | 110.57 | 12.70  | 2.56   | 395.82 | 9.75   | 4.00   | 143.75 |
| 20-20-4-4  | 17.45  | 5.50   | 217.33 | 21.53  | 2.73   | 687.53 | 34.64  | 5.17   | 569.78 |
| 20-20-6-4  | 24.50  | 1.36   | 96.62  | 1.91   | 0.80   | 139.15 | 1.44   | 0.81   | 76.88  |
| 20-20-10-4 | 3.09   | 0.99   | 214.11 | 1.67   | 0.44   | 281.74 | 1.34   | 0.50   | 168.80 |
| 20-20-2-6  | 11.36  | 5.86   | 93.87  | 17.39  | 1.28   | 1257.61 | 8.52  | 3.44  | 147.75 |
| 20-20-4-6  | 24.50  | 10.67  | 129.57 | 28.88  | 4.08   | 608.07 | 27.06  | 8.34  | 224.33 |
| 20-20-6-6  | 43.13  | 14.39  | 199.67 | 80.33  | 10.63  | 656.03 | 57.06  | 16.95  | 236.60 |
| 20-20-8-6  | 50.69  | 24.03  | 110.92 | 137.94 | 34.19  | 303.47 | 69.58  | 32.34  | 115.12 |
| 20-20-10-6 | 2.64   | 0.89   | 196.62 | 2.06   | 0.57   | 277.66 | 1.58   | 0.78   | 101.79 |
| 20-20-2-8  | 12.70  | 5.44   | 133.60 | 18.14  | 2.16   | 741.42 | 8.30   | 4.30   | 93.09  |
| 20-20-4-8  | 45.11  | 20.83  | 116.58 | 134.98 | 32.11  | 320.39 | 40.56  | 36.89  | 9.95   |
| 20-20-6-8  | 79.94  | 43.22  | 84.96  | 1619.81 | 1997.02 | -18.89 | 92.58  | 71.72  | 29.08  |
| 20-20-10-8 | 74.02  | 36.63  | 102.09 | 8493.41 | N/A    | N/A   | 87.84  | 65.69  | 33.73  |
| Average    | 24.54  | 11.30  | 156.10 | 495.07 | 171.47 | 566.83 | 26.17  | 15.61  | 153.95 |

Qi, Jiang, and Shen: Sequential Competitive Facility Location
00(0), pp. 000-000, © 0000
Table 12  Effectiveness of approximation separation on instances with $|I| = |J| = 30$ and varying $p$, $r$-values

| Instance     | SC   | SC-AS | BI   | BI-AS | SCBI  | SCBI-AS |
|--------------|------|-------|------|-------|-------|---------|
|              | Time(s) | Time(s) | Imp(%) | Time(s) | Time(s) | Imp(%) |
| 30-30-3-3    | 34.11  | 25.88  | 31.82 | 8.33  | 2.77  | 201.19  | 6.06  | 4.88  | 24.37 |
| 30-30-6-3    | 450.97 | 414.45 | 8.81  | 21.84 | 10.50 | 108.04  | 19.16 | 19.00 | 0.81  |
| 30-30-9-3    | LIMIT | LIMIT | N/A  | 36.02 | 17.55 | 105.25  | 34.19 | 25.19 | 26.33 |
| 30-30-12-3   | 2198.05 | 2145.91 | 2.43  | 27.42 | 16.16 | 69.73   | 20.20 | 15.33 | 24.13 |
| 30-30-15-3   | 88.42  | 62.84  | 40.70 | 28.59 | 10.80 | 164.83  | 20.49 | 17.22 | 15.94 |
| 30-30-3-6    | 143.83 | 94.08  | 52.88 | 33.52 | 13.42 | 149.72  | 34.53 | 18.81 | 45.52 |
| 30-30-6-6    | 3259.80 | 3196.03 | 2.00  | 140.97 | 54.94 | 156.60  | 130.28 | 107.39 | 17.57 |
| 30-30-9-6    | LIMIT | LIMIT | N/A  | 269.45 | 132.45 | 103.43  | 400.83 | 374.53 | 6.56  |
| 30-30-12-6   | LIMIT | LIMIT | N/A  | 1837.03 | 648.66 | 183.21  | 1981.98 | 2249.34 | -13.49 |
| 30-30-3-9    | 92.06  | 89.02  | 3.42  | 38.48 | 13.64 | 182.14  | 31.13  | 19.89 | 36.09 |
| 30-30-6-9    | 2098.56 | 2294.66 | -8.55 | 314.88 | 85.20 | 269.55  | 304.11 | 237.67 | 21.85 |
| 30-30-3-12   | 55.72  | 42.30  | 31.73 | 25.36 | 4.74  | 435.56  | 20.95  | 16.86 | 19.53 |
| 30-30-6-12   | 2199.31 | 1954.39 | 12.53 | 1026.81 | 353.95 | 190.10  | 965.25  | 1093.69 | -13.31 |
| 30-30-3-15   | 39.92  | 36.64  | 8.96  | 22.73 | 5.91  | 284.87  | 21.39  | 22.58 | -5.55 |
| 30-30-6-15   | 2609.08 | 2430.55 | 7.35  | LIMIT | LIMIT | N/A  | 2234.25 | 2410.95 | -7.91 |
| **Average**  | 969.16 | 941.47 | **16.98** | 153.54 | 52.00 | **201.12** | 143.05 | 143.03 | **17.00** |