A Distributionally Robust Optimization Approach for Two-stage Facility Location Problems

Arash Gourtani[*]   Tri-Dung Nguyen[†]   Huifu Xu[‡]

January 9, 2020

Abstract

In this paper, we consider a facility location problem where customer demand constitutes considerable uncertainty, and where complete information on the distribution of the uncertainty is unavailable. We formulate the optimal decision problem as a two-stage stochastic mixed integer programming problem: an optimal selection of facility locations in the first stage and an optimal decision on the operation of each facility in the second stage. A distributionally robust optimization framework is proposed to hedge risks arising from incomplete information on the distribution of the uncertainty. Specifically, by exploiting the moment information, we construct a set of distributions which contains the true distribution, and where the optimal decision is based on the worst distribution from the set. We then develop two numerical schemes for solving the distributionally robust facility location problem: a semi-infinite programming approach which exploits moments of certain reference random variables, and a semi-definite programming approach which utilizes the mean and correlation of the underlying random variables describing the demand uncertainty. In the semi-infinite programming approach, we apply the well-known linear decision rule approach to the robust dual problem and then approximate the semi-infinite constraints through the conditional value at risk (CVaR) measure. We provide numerical tests to demonstrate the computation and properties of the robust solutions.

Keywords: Distributionally robust optimization, facility location problem, semi-definite programming, semi-infinite programming

---

[*]School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK; email: arash.gourtani@gmail.com.

[†]Mathematical Sciences and Business School (CORMSIS), University of Southampton, Southampton, SO17 1BJ, UK; email: T.D.Nguyen@soton.ac.uk

[‡]Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong; Email: hfxu@se.cuhk.edu.hk

1 Introduction

The classic discrete facility location problem (FLP) involves selecting a subset of facility locations within a finite set of available locations and assigning customers to the selected facilities with the aim to minimize the combined facility setup cost and transportation cost. In the most basic form, the discrete FLPs consist of allocating p facilities to a given list of candidate locations. In these so-called p-median problems the fixed setup cost of all candidate sites are assumed to be equal. The objective function is only to minimize the total service cost to the customers, i.e. the transportation cost. Under the non-homogeneity of the facilities' setup cost, the p-median problem can be extended to uncapacitated facility location problems (UFLP) in which the setup cost is also added to the objective function. The UFLP assumes that facilities can serve an unlimited amount of demand. However, in many practical problems, facilities have capacity constraints and this leads to an important family of FLPs called capacitated facility location problems (CFLP) in which the cheapest-assignment criterion is not sufficient for the optimality of the solution. Note that in the p-median problem, the number of facilities to be installed is fixed to p whereas, in UFLP and CFLP, no such constraints are posed. The p-median, UFLP, and CFLP have been subjects of extensive research and interested readers might refer to Daskin (2011, 2008) and ReVelle and Eiselt (2005) for some comprehensive reviews.

The aforementioned models share certain characteristics such as single-period planning horizon, single product and facility type, and deterministic parameters (i.e. demands, supplies, and costs). However, the deterministic assumption is one of the major drawbacks in coping with many real-world problems. The strategic decisions on facility setup are often capital-intensive, non-repetitive, and span a long time horizon. The decision has to be made at present and hence is subject to risks arising from uncertainties in demands and operations of the established facilities. Hedging the risk, therefore, becomes a vital component of the decision-making process. The facility location problem under uncertainty has attracted considerable attention recently; see, for examples, reviews in Owen and Daskin (1998), Snyder (2006) and Lim et al. (2010). Two major frameworks used to model uncertainty in the facility location problems are stochastic optimization and robust optimization.

In the first framework, stochastic optimization has long been a well-known mathematical method for finding optimal decisions under uncertainty. A key assumption in this approach is that the decision maker has complete information on the distribution of the uncertainty, through either empirical data or subjective judgment. However, in some circumstances, this might turn out to be difficult if not impossible when a strategic decision has to be made well in advance of the realization of the uncertainty.

In the second approach, the robust optimization framework, no assumption is made on the probability distribution of the uncertainty. The traditional proposed measure of robustness is the minmax cost approach in which the cost associated with the worst case scenario is minimized. Some of the examples of the minmax approach can be found in Averbakh and Berman (1997), Averbakh and Berman (2000), Conde (2007) and Snyder and Daskin (2006).

A feasible way to address the issue of distributional uncertainty in stochastic optimization is

to use the available data to construct a set of distributions which contains the true distribution of the uncertainty and make an optimal decision on the basis of the worst distribution from the set. This approach is known as distributionally robust optimization which was proposed by Scarf et al. (1958) and has now been extensively studied over the past few decades. [For some recent developments see Delage and Ye (2010), Esfahani and Kuhn (2018), Wiesemann et al. (2014), Xu et al. (2018), Lu et al. (2015), Santiváñez and Carlo (2018) and references therein.] How to construct the set of distributions depends on the available information and there is no unified framework for this. For examples, if there are some empirical data which allow one to construct a nominal empirical distribution, then one may use the Kantorovich/Wasserstein metric or ϕ-divergence to construct a ball of distributions (Esfahani and Kuhn (2018) and Love and Bayraksan (2015)), if the information comes with statistical quantities such as mean and variance, then one may use moment type conditions (Delage and Ye (2010), Wiesemann et al. (2014) and Xu et al. (2018)). In this paper, we use the moment approach. Specifically, we propose a distributionally robust optimization model for the capacitated facility location problem to deal with future demand uncertainty. We then propose two numerical schemes— namely a semi-definite program and a semi-infinite program— to solve the distributionally robust optimization model. Finally, we provide numerical results for medium size instances taken from the literature.

The remainder of this paper is organized as follows.

(1) In Section 2, we formally describe the deterministic model, a two-stage stochastic model and a distributionally robust formulation of the stochastic model.

(2) We then proceed with discussions on numerical schemes in the following two consecutive sections for solving the robust model depending on the availability of information on the distribution of demands: a semi-definite programming (SDP) scheme in Section 3 and a semi-infinite programming (SIP) scheme in Section 4.

(3) In Section 5, we report numerical test results of the two schemes and draw some conclusions in Section 6.

2   Facility location models

We first introduce the classic deterministic capacitated facility location (D-FLP) problem. By taking into account future uncertainty we extend it to a two-stage stochastic facility location (S-FLP) modelling framework, which then forms the basis for developing a distributionally robust facility location (R-FLP) model.

2.1   Deterministic facility location model

There is a vast body of literature on the deterministic facility location problem. A good review of the related literature is carried out by Owen and Daskin (1998) and Daskin (2011). Suppose

there are up to n facilities to be opened in a set of possible locations $I = \{1, \cdots, n\}$, indexed by i. Let $J = \{1, \cdots, m\}$ be the set of customers, indexed by j. The location decision variable z_i is defined as:

$$z_i = \begin{cases} 1 & \text{if the facility } i \text{ is opened,} \\ 0 & \text{otherwise,} \end{cases}$$

and continuous assignment variable x_{ij} determines the service quantity, or the transportation quantity as we refer to it in this work, assigned from facility i to customer j. Each facility $i \in I$ has a fixed opening cost of b_i and, if opened, has a service capacity s_i that can be used by one or more customers. Likewise, each customer $j \in J$ is characterized by a demand d_j that needs to be satisfied by one or more facilities. In the deterministic framework, it is assumed that the customer demand and the capacity of the facilities are known. The unit service cost of facility i for serving customer j is denoted by c_{ij} and assumed to be known; e.g., transportation costs are proportional to the distances between the facilities and the customers. Note that real-life transportation problems are likely to be unbalanced; i.e. the total demand might exceed the total supply (Goyal (1984)). To avoid the unbalance in the transportation problem, we assume that an external supplier will serve the excess demand w_j of customer j. The unit service cost of this external supply facility, denoted by C_j, is assumed to be sufficiently large, $C_j \geq \max_{i \in I} c_{ij}$, so that the external supply is invoked only when the total supply from all facilities fails to satisfy the demand.

The objective of the facility location problem is to minimize the total fixed cost of opening the facilities and the future transportation costs while satisfying the customer demand and supply capacity constraints. The deterministic FLP is formulated as follows:

$$(\text{D-FLP}) \quad \min_{\mathbf{z}, \mathbf{x}, \mathbf{w}} \quad \sum_i b_i z_i + \sum_{i,j} c_{ij} x_{ij} + \sum_j C_j w_j$$

$$\text{s.t.} \quad \sum_i x_{ij} + w_j \geq d_j, \ \forall \, j \in J, \tag{2.1a}$$

$$\sum_j x_{ij} \leq z_i s_i, \ \forall \, i \in I, \tag{2.1b}$$

$$x_{ij} \geq 0, \ \forall \, j \in J, \ i \in I, \tag{2.1c}$$

$$w_j \geq 0, \ \forall \, j \in J, \tag{2.1d}$$

$$z_i \in \{0, 1\}, \ \forall \, i \in I. \tag{2.1e}$$

Balance constraint (2.1a) ensures that the demands of all customers are met. Constraint (2.1b) prevents the service level assigned to each facility from exceeding its capacity and also ensures that the customers cannot be served by un-built facilities (i.e. when $z_i = 0$, x_{ij} must be equal to zero). Finally, constraints (2.1c)-(2.1e) enforce the non-negativity of the service quantities and the binary nature of the facility allocating decisions, respectively. Variations of deterministic facility location formulations similar to the D-FLP have been well studied in the literature. We refer the readers to Daskin (2011) for more details about these.

2.2 Two-stage stochastic model

The facility location problems involve uncertainties that stem from unpredictability of demand, supply and service costs. Since the facility location decisions are irreversible and capital intensive, it is vital to take into account the future uncertainties when the facility location decisions are made. Louveaux (1986) first introduced a two-stage stochastic program with recourse for solving simple plant location problems and p-median problems where uncertainties in demand, production and transportation costs are considered. Wu et al. (2015) develop a distributionally robust model for the uncapaciated facility location problem under demand uncertainty and derive some nice reformulation of the stochastic model. We follow this approach but extend it to the capacitated cases. There are several differences between our model and that in Wu et al. (2015) which are described below. First of all, Wu et al.'s model is for the un-capacitated case, which means each facility can serve as many customers as it wishes as long as the total transportation distances are minimised as part of the objective function. Their decision variables are which facilities to *fully* serve which customers while our decision variables are to what extend each facility, if opened, serves each customer. When the facilities have unlimited capacity, the demand uncertainty does not play as much of an important role as when there is capacity limitation. This is because, despite uncertain demand, each customer is still being fully served by the nearest facility. In our case, it is possible that the demand exceeds the nearby facilities' capacities and hence we have to buy and transport from one or more expensive sources.

To extend the deterministic model described in subsection 2.1 to a stochastic setting, customer demand is assumed to be stochastic with a known probability distribution. Instead of having a deterministic demand vector $\mathbf{d}$, we use the notation $\mathbf{d}(\boldsymbol{\xi})$ for the stochastic demands that depend on a random vector $\boldsymbol{\xi}$. For convenience in notation, we use $\mathbf{d}(\boldsymbol{\xi})$ and $\boldsymbol{\xi}$ interchangeably; i.e. both $d_j(\boldsymbol{\xi})$ and ξ_j refer to the stochastic demand of customer j. The objective of the two-stage problem is to minimize the sum of fixed investment cost of allocating the facilities and the expected future transportation costs. The resulting mathematical model is given as follows,

$$\text{(S-FLP)} \quad \min_{\mathbf{z}} \quad \sum_i b_i z_i + \mathbb{E}\left[g(\mathbf{z}, \boldsymbol{\xi})\right]$$

$$\text{s.t.} \quad z_i \in \{0, 1\}, \ \forall \, i \in I, \tag{2.2}$$

where $g(\mathbf{z}, \boldsymbol{\xi})$ is the optimal value of the second stage transportation problem

$$g(\mathbf{z}, \boldsymbol{\xi}) := \min_{x, w} \quad \sum_{i,j} c_{ij} x_{ij} + \sum_j C_j w_j,$$

$$\text{s.t.} \quad \sum_i x_{ij} + w_j \geq d_j(\boldsymbol{\xi}), \ \forall \, j \in J, \tag{2.3a}$$

$$\sum_j x_{ij} \leq z_i s_i, \ \forall \, i \in I, \tag{2.3b}$$

$$x_{ij} \geq 0, \ \forall \, i \in I, \ j \in J, \tag{2.3c}$$

$$w_j \geq 0, \ \forall \, j \in J, \tag{2.3d}$$

5

and the expectation is taken w.r.t. the distribution of random vector $\boldsymbol{\xi}$. In this formulation, the decision on $\mathbf{z}$ in the first stage determines the location of new facilities to be built, before the realization of the uncertain demand $\mathbf{d}(\boldsymbol{\xi})$, whilst in the second stage a decision is made to specify allocation of transportation resources after the demand is realized. The optimal value $g(\mathbf{z}, \boldsymbol{\xi})$ of the second-stage problem therefore depends on $\mathbf{z}$ and $\boldsymbol{\xi}$.

2.3 Distributionally robust facility location model

One of the major difficulties which often arises in facility location problems is the lack of complete information on the probability distribution of future customer demands at different locations. We consider a setting where there might be limited information on the probability distribution P of the random parameters $\boldsymbol{\xi}$. Suppose that we are able to construct a set of probability distributions, denoted by $\mathcal{P}$, which contains the true probability distribution P. In order to hedge the risk against ambiguity of the true distribution, we may consider a robust model where the optimal decision on the location of the new facilities to be built and the operation of all facilities is based on the worst distribution from $\mathcal{P}$. In the literature of robust optimization, $\mathcal{P}$ is called the *ambiguity set* which reflects the fact that there is an ambiguity in the true probability distribution in this setup. The corresponding distributionally robust optimization problem can be formulated as

$$(\text{R-FLP}) \qquad \min_{\mathbf{z}} \quad \mathbf{b}^T \mathbf{z} + \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[g(\mathbf{z}, \boldsymbol{\xi}) \right] \tag{2.4}$$
$$\text{s.t.} \quad \mathbf{z} \in \{0, 1\}^n.$$

One of the key elements in distributionally robust optimization is construction of the ambiguity set. Over the past few decades, various statistical methods have been proposed among which the method using moment information of the underlying probability distribution seems to be popular; see Delage and Ye (2010) and references therein. Another widely adopted approach is a mixture distribution which uses a convex combination of some observed and/or predicted distributions. In this paper, we will consider the moment approach which seems to be relevant in the problem setting. We will proceed our discussions through two distinct mathematical formulations: In the first formulation, we assume the mean value and covariance of $\boldsymbol{\xi}$ are known and consequently we reformulate problem (2.4) as a semi-definite program (SDP). In the second formulation, we weaken the assumption by merely assuming the mean value of $\boldsymbol{\xi}$ is known and subsequently reformulate problem (2.4) as a semi-infinite program (SIP). We then develop appropriate numerical procedures for solving the resulting optimization problems.

3 A semi-definite programming approach

In this section, we first formulate the (R-FLP) problem as an integer semi-definite program in Section 3.1. The resulting model, however, has a large number of SDP constraints in addition to the inherent binary variables for the facility location decisions. We propose to resolve the

integrality issue by using a generic genetic algorithm in Section 3.2 and the large number of SDP constraints by using a row generation algorithm in Section 3.3.

3.1 SDP Formulation

We first investigate (R-FLP) with some moment information of $\boldsymbol{\xi}$. Let Ξ denote the range of $\boldsymbol{\xi}$ and $\mathscr{P}(\Xi)$ the set of all probability measures over Ξ. We consider the following ambiguity set,

$$\mathcal{P} = \left\{ P \in \mathscr{P}(\Xi) : \mathbb{E}_P[\boldsymbol{\xi}] = \boldsymbol{\mu}, \quad \mathbb{E}_P[\boldsymbol{\xi}\boldsymbol{\xi}^T] = \mathbf{Q} \right\}, \tag{3.1}$$

where $\boldsymbol{\mu}$ and $\mathbf{Q}$ denote the first and second moments of $\boldsymbol{\xi}$. In practice, the true moments may be unknown. In the literature of distributionally robust optimization, one often specifies a range for these statistical quantities; see for instance Delage and Ye (2010). Assume here that both $\boldsymbol{\mu}$ and $\mathbf{Q}$ can be approximated using empirical data. Quantification of the difference between the ambiguity using true moments and the one using sample average approximation is well documented in Sun and Xu (2015) and Zhang et al. (2016). Its impact on the optimal value and optimal solutions can also be found in these papers.

Let

$$H(\mathbf{z}) = \sup_{P \in \mathcal{P}} \mathbb{E}_P\left[g(\mathbf{z}, \boldsymbol{\xi})\right]. \tag{3.2}$$

Problem (3.2) is related to the classical problem of moments. Here, instead of finding a feasible distribution $P \in \mathcal{P}$, we want to find one which maximizes the expected value of $g(\mathbf{z}, \boldsymbol{\xi})$. For a discussion on the background of the problem of moments, interested readers are referred to Landau (1987). Let $\mathscr{M}^+$ denote the set of all non-negative finite measures on measurable space $(\Xi, \mathscr{B})$. Then

$$
\begin{aligned}
H(\mathbf{z}) := \quad & \sup_{P \in \mathscr{M}^+} \quad \int_{\Xi} g(\mathbf{z}, \boldsymbol{\xi}) P(d\boldsymbol{\xi}) \\
& \text{s.t.} \quad \int_{\Xi} \xi_i \xi_j P(d\boldsymbol{\xi}) = Q_{ij}, \ \forall i, j = 1, \cdots, m, \\
& \qquad \int_{\Xi} \xi_i P(d\boldsymbol{\xi}) = \mu_i, \ \forall i = 1, \cdots, m, \\
& \qquad \int_{\Xi} P(d\boldsymbol{\xi}) = 1,
\end{aligned}
\tag{3.3}
$$

where Q_{ij} denotes the (ij)th component of $\mathbf{Q}$, and μ_j the jth component of $\boldsymbol{\mu}$.

Let $\mathbb{S}^{m \times m}$ denote the space of m by m real matrices. Let $\mathbf{Y} \in \mathbb{S}^{m \times m}$, $\mathbf{y} \in \mathbb{R}^m$ and $y_0 \in \mathbb{R}$ denote the dual variables associated with the moment constraints in (3.3). We can then write

the Lagrangian dual problem of (3.3) as follows,

$$
\begin{aligned}
H_D(\mathbf{z}) := \quad & \min_{\mathbf{Y},\mathbf{y},y_0} \quad \mathbf{Q} \bullet \mathbf{Y} + \boldsymbol{\mu}^{\mathbf{T}}\mathbf{y} + y_0 \\
& \text{s.t.} \quad \boldsymbol{\xi}^T \mathbf{Y} \boldsymbol{\xi} + \boldsymbol{\xi}^T \mathbf{y} + y_0 \geq g(\mathbf{z}, \boldsymbol{\xi}), \ \forall \boldsymbol{\xi} \in \mathbb{R}^m.
\end{aligned}
\tag{3.4}
$$

Here and later on $A \bullet B$ denotes the Frobenius inner product of matrices A and B. This kind of dual formulation is well-known; see for instance Zymler et al. (2013) and (Shapiro et al., 2009, Chapter 6) for general moment problems. It is easy to prove that $H(\mathbf{z}) \leq H_D(\mathbf{z})$; see for instance Bertsimas and Sethuraman (2000).

For strong duality results to hold, we make the following assumption

Assumption 3.1. *The linear matrix inequality* $\mathbf{Q} - \boldsymbol{\mu}\boldsymbol{\mu}^T \succ 0$ *holds, where* $A \succ 0$ *means* A *is positive definite.*

Here and later on we write $A \succeq 0$ for matrix A being positive semi-definite. Note that, for $(\boldsymbol{\mu}, \mathbf{Q})$ to be valid first and second moments of some random variable, it is necessary to have condition $\{\mathbf{Q} - \boldsymbol{\mu}\boldsymbol{\mu}^T \succeq 0\}$ satisfied. Indeed, for any vector v, we have

$$
v^{\mathrm{T}}(\mathbf{Q} - \boldsymbol{\mu}\boldsymbol{\mu}^T)v = v^{\mathrm{T}}\,\mathrm{E}\left[(\boldsymbol{\xi} - \mathrm{E}[\boldsymbol{\xi}])(\boldsymbol{\xi} - \mathrm{E}[\boldsymbol{\xi}])^{\mathrm{T}}\right]v = \mathrm{E}\left[v^{\mathrm{T}}(\boldsymbol{\xi} - \mathrm{E}[\boldsymbol{\xi}])(\boldsymbol{\xi} - \mathrm{E}[\boldsymbol{\xi}])^{\mathrm{T}}v\right]
$$
$$
= \mathrm{E}\left[\left(v^{\mathrm{T}}(\boldsymbol{\xi} - \mathrm{E}[\boldsymbol{\xi}])\right)^2\right] \geq 0,
$$

since $v^{\mathrm{T}}(\boldsymbol{\xi} - \mathrm{E}[\boldsymbol{\xi}])$ is a scalar. Thus, the covariance matrix is positive definite unless the uncertain sources $\boldsymbol{\xi}$ are linearly dependent. Assumption 3.1 is slightly stronger as we replace positive semi-definiteness with positive definiteness, i.e., we explicitly assume that there is no linear dependency among the sources of uncertainty. This is needed for reasons of technicality in proving the strong duality result, which is formally stated in the following proposition.

Proposition 3.1. *Under Assumption 3.1,* $H(\mathbf{z}) = H_D(\mathbf{z})$.

Proof. Let us define $\mathcal{H} = \left\{(M, v) \mid M = M^T, \ M \succ vv^T\right\}$. For any fixed $(M, v) \in \mathcal{H}$, there exists a symmetric matrix W such that $W^2 = (M - vv^T)$. We can then construct a random variable $\boldsymbol{\xi} = WX + v$ such that $\boldsymbol{\xi}$ has a mean value of v and a covariance matrix of $(M - vv^T)$, where $X \in \mathbb{R}^n$ is a random vector which follows the standard multivariate normal distribution. It is easy to verify that ξ satisfies the following:

$$
\int_{\Xi} \xi_i \xi_j P(d\boldsymbol{\xi}) = M_{ij}, \ \forall i, j = 1, \cdots, m,
$$
$$
\int_{\Xi} \xi_i P(d\boldsymbol{\xi}) = v_i, \ \forall i = 1, \cdots, m, \tag{3.5}
$$
$$
\int_{\Xi} P(d\boldsymbol{\xi}) = 1.
$$

In other words, using any choice of $(M, v) \in \mathcal{H}$ to replace $(\mathbf{Q}, \boldsymbol{\mu})$ in the right-hand side of (3.5) would lead to a feasible $\boldsymbol{\xi}$. In addition, we can show that $\mathcal{H}$ is an open set. Therefore, under Assumption 3.1, i.e. $(\mathbf{Q}, \boldsymbol{\mu}) \in \mathcal{H}$, we also have $(\mathbf{Q}, \boldsymbol{\mu})$ to belong to the interior of $\mathcal{H}$. As a result, there exists a neighbourhood $\mathcal{B}$ small enough around $(\mathbf{Q}, \boldsymbol{\mu})$ such that if we replace the right-hand side of (3.5) by any $(M, v) \in \mathcal{B}$, the system of equations (3.5) is still feasible (for some different $\boldsymbol{\xi}$). This is the sufficient condition for having strong duality result to hold; i.e. $H(\mathbf{z}) = H_D(\mathbf{z})$, as stated in (Shapiro, 2001, Proposition 3.4). $\qquad\square$

From the strong duality result, problem (2.4) can be equivalently written as

$$
\begin{aligned}
\min_{\mathbf{z}, \mathbf{Y}, \mathbf{y}, y_0} \quad & \mathbf{b}^T \mathbf{z} + \mathbf{Q} \bullet \mathbf{Y} + \boldsymbol{\mu}^{\mathbf{T}} \mathbf{y} + y_0 \\
\text{s.t.} \quad & \boldsymbol{\xi}^T \mathbf{Y} \boldsymbol{\xi} + \boldsymbol{\xi}^T \mathbf{y} + y_0 \geq g(\mathbf{z}, \boldsymbol{\xi}), \ \forall \boldsymbol{\xi} \in \Xi, \\
& \mathbf{z} \in \{0, 1\}^n.
\end{aligned}
\tag{3.6}
$$

Let us now write down the dual of the transportation problem described in problem (2.3),

$$
\begin{aligned}
g_D(\mathbf{z}, \boldsymbol{\xi}) = \quad \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \quad & \sum_j \alpha_j \xi_j - \sum_i \beta_i (s_i z_i) \\
\text{s.t.} \quad & \alpha_j - \beta_i \leq c_{ij}, \ \forall \, i \in I, \ j \in J, \\
& \alpha_j \leq C_j, \ \forall \, j \in J, \\
& \alpha_j, \beta_i \geq 0, \ \forall \, i \in I, \ j \in J,
\end{aligned}
\tag{3.7}
$$

where α_j, $j \in J$, are dual variables associated with demand constraints (2.3a) and β_i, $i \in I$, are dual variables associated with supply constraints (2.3b). Observe that problem (2.3) satisfies the Mangasarian-Fromovitz constraint qualification (MFCQ). Thus the Lagrange multipliers of the problem are bounded and there exists a positive number C' such that the problem above is equivalent to the following,

$$
\begin{aligned}
\text{(DTP)} \qquad g_D(\mathbf{z}, \boldsymbol{\xi}) = \quad \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \quad & \sum_j \alpha_j \xi_j - \sum_i \beta_i (s_i z_i) \\
\text{s.t.} \quad & \text{constraints of problem (3.7),} \\
& \beta_i \leq C', \forall i \in I.
\end{aligned}
$$

To ease the notation, let $\boldsymbol{\gamma}_\alpha := \boldsymbol{\alpha}$, $\boldsymbol{\gamma}_\beta := \boldsymbol{\beta}$, and $\boldsymbol{\gamma} := (\boldsymbol{\gamma}_\alpha, \boldsymbol{\gamma}_\beta)$. Let Γ denote the feasible set of (DTP). It is easy to observe that Γ is a polyhedral in $\mathbb{R}^{|I|+|J|}$ where $|I|$ and $|J|$ denote the cardinality of the index sets I and J, respectively.

It is easy to observe that Γ is a bounded polyhedral. In addition, the finiteness number of vertices of Γ follows from Balinski and Russakoff (1984).

Let $\{\boldsymbol{\gamma}^1, \cdots, \boldsymbol{\gamma}^N\}$ denote the set of vertices. Using the notation introduced above, we can

rewrite (DTP) in a neater form,

$$g_D(\mathbf{z}, \boldsymbol{\xi}) = \max_{\boldsymbol{\gamma}(\mathbf{z})} \quad \boldsymbol{\xi}^T \boldsymbol{\gamma}_\alpha - \boldsymbol{\gamma}_\beta^T (\mathbf{s} \circ \mathbf{z})$$

$$\text{s.t.} \quad \boldsymbol{\gamma} \in \Gamma, \tag{3.8}$$

where $(\mathbf{s} \circ \mathbf{z})$ denotes an m-dimensional vector with components $s_i z_i$ for $i \in I$. Combining (3.6) and (DTP), we can recast the robust facility location problem (2.4) as a semi-definite program through the following proposition.

Proposition 3.2. *Let $\mathcal{P}$ be defined as in (3.1) and $\Xi = \mathbb{R}^m$. Under Assumption 3.1, the two-stage distributionally robust facility location problem (2.4) can be reformulated as the following semi-definite optimization problem:*

$$(R\text{-}SDP) \quad \min_{\mathbf{z}, \mathbf{Y}, \mathbf{y}, y_0} \quad \mathbf{b}^T \mathbf{z} + \mathbf{Q} \bullet \mathbf{Y} + \boldsymbol{\mu}^\mathbf{T} \mathbf{y} + y_0$$

$$\text{s.t.} \quad \begin{bmatrix} y_0 + \boldsymbol{\gamma}_\beta^T (\mathbf{s} \circ \mathbf{z}) & \frac{1}{2}(\mathbf{y} - \boldsymbol{\gamma}_\alpha)^T \\ \frac{1}{2}(\mathbf{y} - \boldsymbol{\gamma}_\alpha) & \mathbf{Y} \end{bmatrix} \succeq 0, \ \forall \boldsymbol{\gamma} \in \{\boldsymbol{\gamma}^1, \cdots, \boldsymbol{\gamma}^N\} \tag{3.9}$$

$$\mathbf{z} \in \{0, 1\}^n.$$

Proof. It follows from Proposition 3.1 that, under Assumption 3.1, $H(\mathbf{z}) = H_D(\mathbf{z})$ and problem (2.4) can be reformulated as problem (3.6). Note that the reformulation still involves $g(\mathbf{z}, \boldsymbol{\xi})$. Since $g(\mathbf{z}, \boldsymbol{\xi})$ and $g_D(\mathbf{z}, \boldsymbol{\xi})$ are primal and dual LPs of each other and since both of them are feasible (i.e. by setting $\mathbf{x} = 0$, $\omega_j = d_j, \forall j \in J$ in the primal and $\alpha = \beta = 0$ in the dual), strong duality result holds and we have $g(\mathbf{z}, \boldsymbol{\xi}) = g_D(\mathbf{z}, \boldsymbol{\xi})$.

Thus, we can replace the second-stage transportation problem through its dual and rewrite the constraint of problem (3.6) as

$$\boldsymbol{\xi}^T \mathbf{Y} \boldsymbol{\xi} + \boldsymbol{\xi}^T \mathbf{y} + y_0 \geq \max_{\boldsymbol{\gamma} \in \Gamma} \left\{ \boldsymbol{\xi}^T \boldsymbol{\gamma}_\alpha - \boldsymbol{\gamma}_\beta^T (\mathbf{s} \circ \mathbf{z}) \right\}, \quad \forall \boldsymbol{\xi} \in \mathbb{R}^m. \tag{3.10}$$

Since Γ is bounded with a finite set of extreme points $\{\boldsymbol{\gamma}^1, \cdots, \boldsymbol{\gamma}^N\}$, the maximizer of the LP on the R.H.S of (3.10) occurs at one of the extreme points. Thus, (3.10) can be equivalently rewritten as

$$\boldsymbol{\xi}^T \mathbf{Y} \boldsymbol{\xi} + \boldsymbol{\xi}^T \mathbf{y} + y_0 \geq \boldsymbol{\xi}^T \boldsymbol{\gamma}_\alpha - \boldsymbol{\gamma}_\beta^T (\mathbf{s} \circ \mathbf{z}), \quad \forall \boldsymbol{\xi} \in \mathbb{R}^m, \ \forall \boldsymbol{\gamma} \in \{\boldsymbol{\gamma}^1, \cdots, \boldsymbol{\gamma}^N\}.$$

A simple rearrangement yields

$$\min_{\boldsymbol{\xi}} \left\{ \boldsymbol{\xi}^T \mathbf{Y} \boldsymbol{\xi} + \boldsymbol{\xi}^T (\mathbf{y} - \boldsymbol{\gamma}_\alpha) + y_0 + \boldsymbol{\gamma}_\beta^T (\mathbf{s} \circ \mathbf{z}) \right\} \geq 0, \quad \forall \boldsymbol{\gamma} \in \{\boldsymbol{\gamma}^1, \cdots, \boldsymbol{\gamma}^N\}. \tag{3.11}$$

We can show that inequality (3.11) holds if and only if

$$\begin{bmatrix} y_0 + \boldsymbol{\gamma}_\beta{}^T(\mathbf{s} \circ \mathbf{z}) & \frac{1}{2}(\mathbf{y} - \boldsymbol{\gamma}_\alpha)^T \\ \frac{1}{2}(\mathbf{y} - \boldsymbol{\gamma}_\alpha) & \mathbf{Y} \end{bmatrix} \succeq 0, \quad \forall \boldsymbol{\gamma} \in \{\boldsymbol{\gamma}^1, \cdots, \boldsymbol{\gamma}^N\}. \tag{3.12}$$

Here, it is very clear that (3.12) implies (3.11). For the reverse, suppose that (3.12) does not hold; i.e. there exists γ and $(q_0, \boldsymbol{q})$ such that

$$\begin{bmatrix} q_0 & \boldsymbol{q}^T \end{bmatrix} \begin{bmatrix} y_0 + \boldsymbol{\gamma}_\beta{}^T(\mathbf{s} \circ \mathbf{z}) & \frac{1}{2}(\mathbf{y} - \boldsymbol{\gamma}_\alpha)^T \\ \frac{1}{2}(\mathbf{y} - \boldsymbol{\gamma}_\alpha) & \mathbf{Y} \end{bmatrix} \begin{bmatrix} q_0 \\ \boldsymbol{q} \end{bmatrix} < 0.$$

We then need to show that (3.11) does not hold either. If $q_0 \neq 0$, then we can construct $\boldsymbol{\xi} = \boldsymbol{q}/q_0$ and obtain $\boldsymbol{\xi}^T \mathbf{Y} \boldsymbol{\xi} + \boldsymbol{\xi}^T(\mathbf{y} - \boldsymbol{\gamma}_\alpha) + y_0 + \boldsymbol{\gamma}_\beta{}^T(\mathbf{s} \circ \mathbf{z}) < 0$ which means that (3.11) does not hold. If $q_0 = 0$, we have $\boldsymbol{q}^T \mathbf{Y} \boldsymbol{q} < 0$. We can then construct $\boldsymbol{\xi} = \delta \boldsymbol{q}$ with sufficiently large δ such that $\delta^2 \boldsymbol{q}^T \mathbf{Y} \boldsymbol{q} + \delta \boldsymbol{q}^T(\mathbf{y} - \boldsymbol{\gamma}_\alpha) + y_0 + \boldsymbol{\gamma}_\beta{}^T(\mathbf{s} \circ \mathbf{z}) < 0$ which also means that (3.11) does not hold.

Finally, we can replace the constraint in (3.4) with the SDP constraint (3.12) and obtain the SDP (3.9).

□

Notice that the derivation from inequality (3.11) to inequality (3.12) requires $\Xi \equiv \mathbb{R}^m$. In practice, there is often some information on the bounds of the uncertain parameters. For example the customer demand cannot take negative values. In order to handle the indefinite set of constraints that appears in problem (3.6) for this case, we will approximate the indefinite constraint with a finite set of semi-definite constraints as shown next.

Suppose the support set is specified as $\Xi = \prod_{j \in J} \Xi_j$, where $\Xi_j = [\underline{\xi}_j, \bar{\xi}_j]$ for all $j \in J$ and $\underline{\boldsymbol{\xi}}$ and $\bar{\boldsymbol{\xi}}$ are some given lower and upper bounds. For reasons of technicality in proving strong duality, we make an assumption that there exists a random vector X with support set Ξ such that $E[X] = 0$ and $E[XX^T] = I$, where $I \in \mathbb{R}^{m \times m}$ is the identity matrix. In practice, the support set of ξ may not be necessarily box structured. In such a case, we might consider a box within the support set of ξ such that the probability of ξ taking values outside the box is negligible. It is possible to investigate the difference between the ambiguity sets before and after cut and its impact on the optimal value and optimal decisions although we have not attempted to do so in this paper.

Under this new assumption and Assumption 3.1, we can show that the strong duality result still holds and the proof is very similar to that of Proposition 3.1. The only difference is in the way that we construct the random variable X (i.e. instead of choosing a multi-variate normal random variable, we choose a random vector X such that $E[X] = 0$ and $E[XX^T] = I$). It is noted that the new assumption can be relaxed further by a proper scaling of the random variables $\boldsymbol{\xi}$.

Once we have derived the strong duality result, problem (3.4) becomes

$$
\begin{aligned}
H_D(\mathbf{z}) = \min_{\mathbf{Y}, \mathbf{y}, y_0} \quad & \mathbf{Q} \bullet \mathbf{Y} + \boldsymbol{\mu}^{\mathbf{T}} \mathbf{y} + y_0 \\
\text{s.t.} \quad & \boldsymbol{\xi}^T \mathbf{Y} \boldsymbol{\xi} + \boldsymbol{\xi}^T \mathbf{y} + y_0 \geq g(\mathbf{z}, \boldsymbol{\xi}), \ \forall \boldsymbol{\xi} \in \Xi,
\end{aligned}
\tag{3.13}
$$

where the only difference compared to problem (3.4) is that we have replaced the semi-infinite constraint $\{\forall \boldsymbol{\xi} \in \mathbb{R}^n\}$ with $\{\forall \boldsymbol{\xi} \in \Xi\}$. Inequality (3.11) now become

$$
\min_{\underline{\boldsymbol{\xi}} \leq \boldsymbol{\xi} \leq \bar{\boldsymbol{\xi}}} \left[\boldsymbol{\xi}^T \mathbf{Y} \boldsymbol{\xi} + \boldsymbol{\xi}^T (\mathbf{y} - \boldsymbol{\gamma}_\alpha) + y_0 + \boldsymbol{\gamma}_\beta{}^T (\mathbf{s} \circ \mathbf{z}) \right] \geq 0, \quad \forall \boldsymbol{\gamma} \in \{\boldsymbol{\gamma}^1, \cdots, \boldsymbol{\gamma}^N\},
\tag{3.14}
$$

which is equivalent to

$$
\phi(\boldsymbol{\gamma}) \geq 0, \ \forall \boldsymbol{\gamma} \in \{\boldsymbol{\gamma}^1, \cdots, \boldsymbol{\gamma}^N\},
\tag{3.15}
$$

where $\phi(\boldsymbol{\gamma})$ is the optimal value of the following program

$$
\min_{\boldsymbol{\xi}} \left\{ \begin{bmatrix} 1 & \boldsymbol{\xi} \end{bmatrix} \begin{bmatrix} y_0 + \boldsymbol{\gamma}_\beta{}^T (\mathbf{s} \circ \mathbf{z}) & \frac{1}{2}(\mathbf{y} - \boldsymbol{\gamma}_\alpha)^T \\ \frac{1}{2}(\mathbf{y} - \boldsymbol{\gamma}_\alpha) & \mathbf{Y} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{\xi} \end{bmatrix} \right\}
$$

$$
\text{s.t.} \quad \begin{bmatrix} 1 & \boldsymbol{\xi} \end{bmatrix} V_j \begin{bmatrix} 1 \\ \boldsymbol{\xi} \end{bmatrix} \leq 0, \qquad \forall j \in J,
$$

and where $V_j = \begin{bmatrix} \bar{\xi}_j \underline{\xi}_j & v_j^T \\ v_j & I_j \end{bmatrix}$, with I_j denoting an $m \times m$ matrix with all elements being 0 except 1 at (j, j), and v_j is an m-dimensional vector with all components are equal to zero except for the jth element which is equal to $-(\underline{\xi}_j + \bar{\xi}_j)/2$. The S-lemma (Derinkuyu and Pınar (2006)) provides a sufficient condition for the non-negativity of the quadratic objective function over the quadratic inequalities corresponding to the bounds. In other words, for the conditions (3.15) to be satisfied for each $\boldsymbol{\gamma} \in \Gamma$, it suffices that there exists $\mathbf{h} \geq 0$ such that

$$
\begin{bmatrix} y_0 + \boldsymbol{\gamma}_\beta{}^T (\mathbf{s} \circ \mathbf{z}) & \frac{1}{2}(\mathbf{y} - \boldsymbol{\gamma}_\alpha)^T \\ \frac{1}{2}(\mathbf{y} - \boldsymbol{\gamma}_\alpha) & \mathbf{Y} \end{bmatrix} + \sum_{j \in J} h_j V_j \succeq 0,
$$

where h_j denotes the jth component of vector $\mathbf{h}$. Consequently, the SDP problem (3.9) can be

reformulated as

$$
\min_{\mathbf{z},\mathbf{Y},\mathbf{y},y_0,\mathbf{h}} \quad \mathbf{b}^T\mathbf{z} + \mathbf{Q}\bullet\mathbf{Y} + \boldsymbol{\mu}^{\mathbf{T}}\mathbf{y} + y_0
$$

$$
\text{s.t.} \quad \begin{bmatrix} y_0 + \boldsymbol{\gamma_\beta}^T(\mathbf{s}\circ\mathbf{z}) & \frac{1}{2}(\mathbf{y}-\boldsymbol{\gamma_\alpha})^T \\[2mm] \frac{1}{2}(\mathbf{y}-\boldsymbol{\gamma_\alpha}) & \mathbf{Y} \end{bmatrix} + \sum_{j\in J} h_j V_j \succeq 0, \ \ \forall\boldsymbol{\gamma}\in\{\boldsymbol{\gamma}^1,\cdots,\boldsymbol{\gamma}^N\}, \tag{3.16}
$$

$$
\mathbf{h} \geq 0,
$$

$$
\mathbf{z} \in \{0,1\}^n.
$$

Remark: Problem (3.16) is very similar to problem (3.9) except for the newly introduced decision variable $\mathbf{h}$. If we restrict $\mathbf{h}=0$, then problem (3.16) is exactly the same as problem (3.9). Each $\mathbf{h}>0$ essentially enlarges the feasible domain of $(\mathbf{Y},\mathbf{y},y_0)$ in problem (3.9) to a larger feasible domain of $(\mathbf{Y},\mathbf{y},y_0)$ in problem (3.16).

3.2   Genetic algorithm for finding z

From a computational perspective, problems (3.9) and (3.16) are complex to solve due to both the presence of the binary variable $\mathbf{z}$ and the potentially large number of SDP constraints. The issue of having binary variables $\mathbf{z}$ is unavoidable as this is an intrinsic part of the facility location problems, even in the deterministic case. For practical purposes, we utilize a genetic algorithm (GA) to search for the optimal facility location variables. To this end, we rewrite Problem (3.16) in a compact form on decision variable $\mathbf{z}$ as follows,

$$
\min_{\mathbf{z}\in\{0,1\}^n} \quad \mathbf{b}^{\mathbf{T}}\mathbf{z} + H_D(\mathbf{z}), \tag{3.17}
$$

where $H_D(\mathbf{z})$ is defined in Formulation (3.4) and has a SDP reformulation similar to Formulation (3.16) except that $\mathbf{z}$ is not a decision variable and the fixed term $\mathbf{b}^{\mathbf{T}}\mathbf{z}$ is taken out of the objective function.

As long as we are able to evaluate the fitness function $\mathbf{b}^{\mathbf{T}}\mathbf{z}+H_D(\mathbf{z})$ for each $\mathbf{z}$, the optimisation problem (3.17) can be embedded in a generic genetic algorithm[1]. In this case, we view the decision variable $\mathbf{z}$ as genomes and the genetic algorithm keeps updating them through operations such as mutation and cross-over to find better solutions. The genetic algorithm is a probabilistic search method that mimics the biological model of natural selection. It applies the principle of "survival of the fittest" to a population of potential solutions to produce progressively better solutions over the generations. We refer the interested readers to Davis (1991) for more details about the genetic algorithm.

---

[1]Other local search technique can be used too.

3.3   Row generation algorithm for evaluating the fitness function

The fitness function is the optimal value of a SDP problem of a similar form as in (3.16) with N semi-definite constraints, where N is the number of vertices of the DTP polyhedra Γ, that is the feasible space of the DTP. The challenge here is that N grows exponentially as the problem size (n, m) increases (Balinski and Russakoff (1984)). We resolve this issue by utilizing a row generation (RG) algorithm. The RG algorithm starts by including only a subset of constraints and solves the restricted problem. The optimal solution obtained might violate some constraints of the original problem. The next step is to identify such violated constraints which are then added to the restricted problem in an iterative manner. This procedure is applied until there is no further violating constraint. In that case, the optimal solution of the restricted problem is also the optimal solution of the original problem.

A key component of a row generation algorithm is to identify violating constraints. Suppose, at iteration k, we solve problem (3.16) with the subset of SDP constraints and obtain $\left(\mathbf{Y}^{(k)}, \mathbf{y}^{(k)}, y_0^{(k)}, \mathbf{h}^{(k)} \right)$. To find a violating constraint, we need to find $\boldsymbol{\gamma} \in \Gamma$ such that

$$\min_{\underline{\boldsymbol{\xi}} \, \leq \, \boldsymbol{\xi} \, \leq \, \bar{\boldsymbol{\xi}}} \left[\boldsymbol{\xi}^T \mathbf{Y} \boldsymbol{\xi} + \boldsymbol{\xi}^T (\mathbf{y} - \boldsymbol{\gamma}_\alpha) + y_0 + \boldsymbol{\gamma}_\beta^T (\mathbf{s} \circ \mathbf{z}) \right] < 0,$$

since this is the complementary condition of the feasibility constraint (3.14).

This can be done by solving

$$\min_{\boldsymbol{\gamma} \in \Gamma, \underline{\boldsymbol{\xi}} \, \leq \, \boldsymbol{\xi} \, \leq \, \bar{\boldsymbol{\xi}}} \left[\boldsymbol{\xi}^T \mathbf{Y}^{(k)} \boldsymbol{\xi} + \boldsymbol{\xi}^T (\mathbf{y}^{(k)} - \boldsymbol{\gamma}_\alpha) + y_0^{(k)} + \boldsymbol{\gamma}_\beta^T (\mathbf{s} \circ \mathbf{z}) \right] \tag{3.18}$$

and checking whether the optimal value is less than zero, in which case we have identified a violating constraint. Otherwise, we conclude that the relaxed solution $\left(\mathbf{Y}^{(k)}, \mathbf{y}^{(k)}, y_0^{(k)}, \mathbf{h}^{(k)} \right)$ is an optimal solution of the original SDP problem. Problem (3.18) is a non-convex quadratic optimisation problem; i.e., still involving a bilinear term in the objective function between decision variable ξ and γ. Despite the NP-hardness nature of the problem, there are several numerical schemes for solving practical problems globally such as in Chen and Burer (2012), Bonami et al. (2018), Gondzio and Yildirim (2018) and Xia et al. (2019). In our numerical scheme, we develop a simple iterative approach to alternatively fix γ to solve the (convex) quadratic program on ξ and then fix the newly found ξ to solve the linear program on γ. While this approach does not provide us a definite conclusion on the feasibility of $\left(\mathbf{Y}^{(k)}, \mathbf{y}^{(k)}, y_0^{(k)}, \mathbf{h}^{(k)} \right)$ if the local optimal (ξ, γ) results in a non-negative objective value, it is still effective in identifying violating constraints when this results in a negative objective value. Here, we note that a global optimisation procedure is only needed in the very last iteration of the row generation algorithm to verify the feasibility of the relaxed optimal solution.

4 A semi-infinite approach

In this section we consider the case when the uncertainty set is defined through the first moment[2]

$$\mathcal{P} = \{P \in \mathscr{P} : \mathbb{E}_P[\boldsymbol{\xi}] = \boldsymbol{\mu}\}, \tag{4.1}$$

where $\boldsymbol{\mu}$ is the true mean value of the random demand $\boldsymbol{\xi}$.

We first formulate the (R-FLP) problem as an integer semi-infinite program (SIP) in Section 4.1. Similar to the SDP formulation, the resulting model still has the inherent binary variables for the facility location decisions and hence we use the same genetic algorithm as described in Section 3. The major additional challenge is that the resulting model has an infinite number of constraints. We first utilise a linear decision rule approximation approach in Section 4.2 to simplify the SIP model by restricting the second-stage decision variables as linear functions of the uncertainties. We then utilise the conditional value at risk approximations over the sets of infinite constraints. Finally, we use sample approximation to approximate the semi-infinite programs, i.e., both the original SIP and the linear-decision rule approximation, and the CVaR formulation in Section 4.4.

4.1 Semi-infinite Formulation

Let us reconsider the inner maximization problem associated with the robust problem (2.4)

$$H(\mathbf{z}) := \sup_{P \in \mathcal{P}} \mathbb{E}_P\left[g(\mathbf{z}, \boldsymbol{\xi})\right].$$

We can derive the dual formulation with respect to the moment condition as

$$
\begin{aligned}
H_D(\mathbf{z}) = \quad &\min_{\lambda_0, \boldsymbol{\lambda}} \quad \lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{\mu} \\
&\text{s.t.} \quad g(\mathbf{z}, \boldsymbol{\xi}) \le \lambda_0 + \boldsymbol{\xi}^T \boldsymbol{\lambda}, \; \forall \boldsymbol{\xi} \in \Xi,
\end{aligned}
\tag{4.2}
$$

where $\Xi \subset \mathbb{R}^m$ is the support set of $\boldsymbol{\xi}$, $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $\lambda_0 \in \mathbb{R}$ are the dual variables associated with the moment constraints and the normalization constraint, respectively. Since the support set Ξ is infinite, the dual problem (4.2) is a linear semi-infinite programming problem (SIP). It is important to note that (4.2) is a deterministic semi-infinite programming problem. If Ξ is structured, e.g., polyhedral or semi-algebraic and g is linear or quadratic w.r.t. $\boldsymbol{\xi}$, then through the well known S-lemma, the semi-infinite system can be represented as an SDP; see for instance Zymler et al. (2013). Here, we don't assume any special structure as such. To avoid duality gap, we assume that the regularity conditions specified in Shapiro et al. (2009) hold. Specifically, we assume that the dual problem has a non-empty and bounded set of optimal solutions and also the support set Ξ is convex and compact. The second-stage maximization problem in (2.4) can

---

[2]It is possible to include the second moment of $\boldsymbol{\xi}$ and the mathematical derivation on strong duality results still holds as shown in Section 3. The numerical scheme to follow would still be applicable (except the problem size is larger). We exclude the second moment in this work for clarity.

be replaced by its dual as follows,

$$\text{(R-SIP)} \quad \min_{\mathbf{z}, \lambda_0, \boldsymbol{\lambda}} \quad \mathbf{b}^T \mathbf{z} + \lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{\mu}$$

$$\text{s.t.} \quad g(\mathbf{z}, \boldsymbol{\xi}) \leq \lambda_0 + \boldsymbol{\xi}^T \boldsymbol{\lambda}, \ \forall \boldsymbol{\xi} \in \Xi, \tag{4.3}$$

$$\mathbf{z} \in \{0,1\}^n,$$

or equivalently

$$\min_{\mathbf{z}, \mathbf{x}(\cdot), \mathbf{w}(\cdot), \lambda_0, \boldsymbol{\lambda}} \quad \mathbf{b}^T \mathbf{z} + \lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{\mu}$$

$$\text{s.t.} \quad \mathbf{c} \bullet \mathbf{x}(\boldsymbol{\xi}) + \mathbf{C} \bullet \mathbf{w}(\boldsymbol{\xi}) \leq \lambda_0 + \boldsymbol{\xi}^T \boldsymbol{\lambda}, \ \forall \boldsymbol{\xi} \in \Xi,$$

$$\left(\mathbf{x}(\boldsymbol{\xi}), \mathbf{w}(\boldsymbol{\xi}) \right) \in \mathcal{G}(\mathbf{z}, \boldsymbol{\xi}), \ \forall \boldsymbol{\xi} \in \Xi, \tag{4.4}$$

$$\mathbf{z} \in \{0,1\}^n,$$

where $\mathbf{x}(\boldsymbol{\xi}) \in \mathbb{R}^{n \times m}$ and $\mathbf{w}(\boldsymbol{\xi}) \in \mathbb{R}^m$ are optimal transportation decisions for each fixed $\mathbf{z}$ and for each realization of $\boldsymbol{\xi}$, and where $\mathbf{c}$ is the matrix of transportation cost coefficients and $\mathbf{C}$ is the cost vector of serving customers from the external source. Moreover, $\mathcal{G}(\mathbf{z}, \boldsymbol{\xi})$ is the feasible regions associated with the second-stage problem (2.3).

4.2 Linear decision rule approximation

One of the main challenges in solving the semi-infinite problem above is the dependence of the second-stage transportation variables $\mathbf{x}(\boldsymbol{\xi})$ and $\mathbf{w}(\boldsymbol{\xi})$ on random variable $\boldsymbol{\xi}$. These "adjustable" variables are often referred to as *decision rules*, and their presence could often complicate the solution procedure. Formally, a decision rule $\mathbf{x}(\boldsymbol{\xi})$ can be defined as a vector-valued function, mapping the random variables $\boldsymbol{\xi} \in \mathbb{R}^m$ with support set Ξ into the decisions. The decision rule problem can be interpreted as identifying the best decision $\mathbf{x}(\boldsymbol{\xi}) \in \Delta \subset \mathcal{X}$ once $\boldsymbol{\xi}$ is observed, where $\mathcal{X}$ denotes the set of all the mapping from Ξ to $\mathbb{R}^{n \times m}$ and Δ a subset of $\mathcal{X}$.

A tractable approximation scheme to deal with the decision rules is to restrict their feasible set to the ones that have a functionality affine relation with the uncertain random variables (that are affine functions of the uncertain data). This approach was proposed by Ben-Tal et al. (2004) and was extended in Shapiro and Nemirovski (2005) and Kuhn et al. (2011) to develop tractable numerical procedure for stochastic programming problems. Here, we take the initiative to apply the linear decision rule (LDR) approximation to problem (4.4); that is, we impose the dependence of transportation decisions on the random demand to follow linear functions

$$\mathbf{x}(\boldsymbol{\xi}) = \mathbf{X}\boldsymbol{\xi} + \mathbf{x^0},$$
$$\mathbf{w}(\boldsymbol{\xi}) = \mathbf{W}\boldsymbol{\xi} + \mathbf{w^0},$$

where $\mathbf{X} \in \mathbb{R}^{(nm \times m)}, \mathbf{W} \in \mathbb{R}^{m \times m} \in \mathbb{R}^{n \times m}, \mathbf{x^0} \in \mathbb{R}^{n \times m},$ and $\mathbf{w^0} \in \mathbb{R}^m$. Consequently, problem (4.4) can be approximated as

$$
\begin{aligned}
\text{(R-LDR)} \quad \min_{\mathbf{z}, \mathbf{X}, \mathbf{x^0}, \mathbf{W}, \mathbf{w^0}, \lambda_0, \boldsymbol{\lambda}} \quad & \mathbf{b}^T \mathbf{z} + \lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{\mu} \\
\text{s.t.} \quad & \mathbf{c} \bullet (\mathbf{X}\boldsymbol{\xi}) + \mathbf{x^0} + \mathbf{C} \bullet (\mathbf{W}\boldsymbol{\xi}) + \mathbf{w^0} \leq \lambda_0 + \boldsymbol{\xi}^T \boldsymbol{\lambda}, \ \forall \boldsymbol{\xi} \in \Xi, \\
& \left(\mathbf{X}\boldsymbol{\xi} + \mathbf{x^0}, \mathbf{W}\boldsymbol{\xi} + \mathbf{w^0} \right) \in \mathcal{G}(\mathbf{z}, \boldsymbol{\xi}), \ \forall \boldsymbol{\xi} \in \Xi, \\
& \mathbf{z} \in \{0,1\}^n.
\end{aligned} \tag{4.5}
$$

Note that here we are slightly abusing the notation in this formulation: $\boldsymbol{\xi}$ should be understood as a parameter rather than a random variable. Indeed, it represents a realization of the random vector $\boldsymbol{\xi}$. The optimal value of the LDR approximation problem will generate an upper bound on the optimal value of original robust problem (4.4).

4.3   Conditional value at risk approximation

Having defined the LDR formulation of the original robust semi-infinite problem, we approximate the first semi-infinite constraint with Conditional Value at Risk (CVaR) and then approximate the latter through Monte Carlo sampling to reduce the number of constraints. One of the main advantages of using CVaR is that it converts the semi-infinite number of constraints into a single constraint. A recent study by Anderson et al. (2014) has shown promising performance of CVaR approximation in dealing with semi-infinite problems. The CVaR approximation method has been extensively used in stochastic programming for approximating the chance constraints and we refer the readers to Sun et al. (2012) and Hong et al. (2011) for more details.

In the case of our LDR problem, we start by considering CVaR approximation of the first semi-infinite constraints in problem (4.5). To ease the notation, let

$$
h(\mathbf{X}, \mathbf{x^0}, \mathbf{W}, \mathbf{w^0}, \lambda_0, \boldsymbol{\lambda}, \boldsymbol{\xi}) := \mathbf{c} \bullet (\mathbf{X}\boldsymbol{\xi}) + \mathbf{x^0} + \mathbf{C} \bullet (\mathbf{W}\boldsymbol{\xi}) + \mathbf{w^0} - \lambda_0 - \boldsymbol{\xi}^T \boldsymbol{\lambda},
$$

and $\mathcal{Q} := (\mathbf{X}, \mathbf{x^0}, \mathbf{W}, \mathbf{w^0}, \lambda_0, \boldsymbol{\lambda})$. The semi-infinite constraint of (4.5) can be written as

$$
\sup_{\boldsymbol{\xi} \in \Xi} h(\mathcal{Q}, \boldsymbol{\xi}) \leq 0.
$$

Let $\tilde{\boldsymbol{\xi}}$ be a continuous random vector with support Ξ. Then $\sup_{\boldsymbol{\xi} \in \Xi} h(\mathcal{Q}, \boldsymbol{\xi})$ can be approximated by CVaR of $h(\mathcal{Q}, \tilde{\boldsymbol{\xi}})$, which is defined as

$$
\text{CVaR}_\beta(h(\mathcal{Q}, \tilde{\boldsymbol{\xi}})) := \min_{\eta \in \mathbb{R}} \Phi_\beta(\mathcal{Q}, \eta),
$$

where

$$\Phi_\beta(\mathcal{Q}, \eta) := \eta + \frac{1}{1-\beta} \int_{\tilde{\boldsymbol{\xi}} \in \Xi} \left(h(\mathcal{Q}, \tilde{\boldsymbol{\xi}}) - \eta \right)_+ \tilde{P}(d\tilde{\boldsymbol{\xi}}),$$

$(\tau)_+ = \max(0, \tau)$, and $\tilde{P}$ denotes the distribution of $\tilde{\boldsymbol{\xi}}$. Consequently, the CVaR approximation of the semi-infinite constraint in problem (4.5) can be expressed as follows:

$$\min_{\eta \in \mathbb{R}} \left(\eta + \frac{1}{1-\beta} \mathbb{E}\left[(h(\mathcal{Q}, \tilde{\boldsymbol{\xi}}) - \eta)_+ \right] \right) \leq 0. \tag{4.6}$$

It is important to distinguish the expectation $\mathbb{E}[\cdot]$ here from the expectation $\mathbb{E}[\cdot]$ in (S-FLP). The former should be understood as a mathematical expectation taken w.r.t. any distribution of any random variable $\tilde{\boldsymbol{\xi}}$ with support set Ξ. In other words, here, the $\tilde{\boldsymbol{\xi}}$ does not have to be identical to the $\boldsymbol{\xi}$ in (S-FLP). For example, we may set $\tilde{\boldsymbol{\xi}}$ as a random variable with uniform distribution over Ξ. Of course, the selection of $\tilde{\boldsymbol{\xi}}$ and its distribution will affect the quantity of CVaR of h and the rate of approximation to its essential supremum.

Under some mild conditions, we can show that the error arising from the approximation scheme does not have significant impact on the optimal value; see Anderson et al. (2014). By replacing the constraint in the original LDR problem, we can write the CVaR approximation problem as

$$\text{(R-CVaR)} \quad \min_{\eta, \mathbf{z}, \mathbf{X}, \mathbf{x^0}, \mathbf{W}, \mathbf{w^0}, \lambda_0, \boldsymbol{\lambda}} \quad \mathbf{b}^T \mathbf{z} + \lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{\mu}$$

$$\text{s.t.} \quad \eta + \frac{1}{1-\beta} \mathbb{E}\left[\left(h(\mathbf{X}, \mathbf{x^0}, \mathbf{W}, \mathbf{w^0}, \lambda_0, \boldsymbol{\lambda}, \tilde{\boldsymbol{\xi}}) - \eta \right)_+ \right] \leq 0,$$

$$\left(\mathbf{X}\boldsymbol{\xi} + \mathbf{x^0}, \mathbf{W}\boldsymbol{\xi} + \mathbf{w^0} \right) \in \mathcal{G}(\mathbf{z}, \boldsymbol{\xi}), \ \forall \boldsymbol{\xi} \in \Xi, \tag{4.7}$$

$$\mathbf{z} \in \{0, 1\}^n.$$

Under the Slater constraint qualification of the LDR problem (4.5), we can demonstrate, similar to (Anderson et al., 2014, Theorem 4), that the optimal solution of the CVaR approximation problem (4.7) converges to optimal solution of the LDR problem as $\beta \to 1$.

4.4 Discretization through sampling

One of the well-known solution approaches for semi-infinite programs is random discretization. The basic idea is to construct a tractable sub-problem by considering a randomly drawn finite subset of constraints and hence enlarging the solution set. Calafiore and Campi (2005, 2006) investigated this approach and used Monte Carlo sampling (often referred to as sample average approximation) to approximate the convex problems consisting of linear objectives and semi-infinite constraints. They showed that the resulting randomized solution fails to satisfy only a small proportion of the original constraints for a sufficiently large sample size. An explicit bound

on the measures of the original constraints that may be violated by the randomized solution is derived. The approach has been shown numerically efficient and it has been widely applied to various stochastic and robust programs. We refer interested readers to Campi and Garatti (2011), Shapiro (2003) and references therein.

In this paper, we apply the Monte Carlo sampling approach respectively to the original semi-infinite problem (4.4), its LDR approximation (4.5) and the CVaR formulation (4.7).

Let $\mathcal{K} = \{1, \cdots, K\}$ denote the finite set of sample indices and $\boldsymbol{\xi}^1, \cdots, \boldsymbol{\xi}^K$ an independent and identically distributed (i.i.d) sampling of $\boldsymbol{\xi}$. We may construct a discretized approximation of problem (4.4) through Monte Carlo sampling as follows,

$$
\begin{aligned}
\min_{\mathbf{z}, x(\xi^k), w(\xi^k): k \in \mathcal{K}, \lambda_0, \boldsymbol{\lambda}} \quad & \mathbf{b}^T \mathbf{z} + \lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{\mu} \\
\text{s.t.} \quad & \mathbf{c} \bullet \mathbf{x}(\boldsymbol{\xi}^k) + \mathbf{C} \bullet \mathbf{w}(\boldsymbol{\xi}^k) \le \lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{\xi}^k, \ \forall k \in \mathcal{K}, \\
& \left(\mathbf{x}(\boldsymbol{\xi}^k), \mathbf{w}(\boldsymbol{\xi}^k) \right) \in \mathcal{G}(\mathbf{z}, \boldsymbol{\xi}^k), \ \forall k \in \mathcal{K}, \\
& \mathbf{z} \in \{0, 1\}^n.
\end{aligned}
\tag{4.8}
$$

This kind of discretization scheme was recently applied to a distributionally robust formulation of a two-stage stochastic unit commitment problem. From a mathematical perspective, it is justified in that under some mild conditions, one can show the convergence of the optimal value of the discretized problem to its true counterpart almost surely as the sample size increases, see details in (Xu et al., 2018, Section 3.1).

Similarly the discretized LDR problem (4.5) can be formulated as

$$
\begin{aligned}
\min_{\mathbf{z}, \mathbf{X}, \mathbf{x^0}, \mathbf{W}, \mathbf{w^0}, \lambda_0, \boldsymbol{\lambda}} \quad & \mathbf{b}^T \mathbf{z} + \lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{\mu} \\
\text{s.t.} \quad & \mathbf{c} \bullet (\mathbf{X} \boldsymbol{\xi}^k) + \mathbf{x^0} + \mathbf{C} \bullet (\mathbf{W} \boldsymbol{\xi}^k) + \mathbf{w^0} \le \lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{\xi}^k, \ \forall k \in \mathcal{K}, \\
& \left(\mathbf{X} \boldsymbol{\xi}^k + \mathbf{x^0}, \mathbf{W} \boldsymbol{\xi}^k + \mathbf{w^0} \right) \in \mathcal{G}(\mathbf{z}, \boldsymbol{\xi}^k), \ \forall k \in \mathcal{K}, \\
& \mathbf{z} \in \{0, 1\}^n,
\end{aligned}
\tag{4.9}
$$

and, finally, we apply the sample average approximation (SAA) scheme to CVaR approximation problem (4.7) as follows,

$$
\begin{aligned}
\min_{\eta, \mathbf{z}, \mathbf{X}, \mathbf{x^0}, \mathbf{W}, \mathbf{w^0}, \lambda_0, \boldsymbol{\lambda}} \quad & \mathbf{b}^T \mathbf{z} + \lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{\mu} \\
\text{s.t.} \quad & \eta + \frac{1}{(1 - \beta) K} \sum_{k=1}^{K} \left[\left(h(\mathbf{X}, \mathbf{x^0}, \mathbf{W}, \mathbf{w^0}, \lambda_0, \boldsymbol{\lambda}, \boldsymbol{\xi}^k) - \eta \right)_+ \right] \le 0, \\
& \left(\mathbf{X} \boldsymbol{\xi}^k + \mathbf{x^0}, \mathbf{W} \boldsymbol{\xi}^k + \mathbf{w^0} \right) \in \mathcal{G}(\mathbf{z}, \boldsymbol{\xi}^k), \ \forall k \in \mathcal{K}, \\
& \mathbf{z} \in \{0, 1\}^n.
\end{aligned}
\tag{4.10}
$$

Compared to (4.8), the CVaR approximation scheme allows one to take a few samples at the tail rather than essential superemum of h, and in that way smooth up or stabilize the numerical computation; see Anderson et al. (2014). In the case of CVaR formulation, we replace the CVaR constraint with the equivalent system of linear inequalities by introducing additional positive dummy variables $\theta^1, \cdots \theta^k$ as follows

$$
\begin{cases}
\eta + \dfrac{1}{(1-\beta)K} \displaystyle\sum_{k=1}^{K} \theta^k \leq 0, \\[4mm]
h(\mathbf{X}, \mathbf{x^0}, \mathbf{W}, \mathbf{w^0}, \lambda_0, \boldsymbol{\lambda}, \boldsymbol{\xi}^k) - \eta \leq \theta^k, \ \forall k \in \mathcal{K}, \\[3mm]
\theta^k \geq 0, \ \forall k \in \mathcal{K}.
\end{cases}
\tag{4.11}
$$

The substitution results in

$$
\begin{aligned}
\min_{\theta, \eta, \mathbf{z}, \mathbf{X}, \mathbf{x^0}, \mathbf{W}, \mathbf{w^0}, \lambda_0, \boldsymbol{\lambda}} \quad & \mathbf{b}^T \mathbf{z} + \lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{\mu} \\
\text{s.t.} \quad\quad & \eta + \frac{1}{(1-\beta)K} \sum_{k=1}^{K} \theta^k \leq 0, \\
& h(\mathbf{X}, \mathbf{x^0}, \mathbf{W}, \mathbf{w^0}, \lambda_0, \boldsymbol{\lambda}, \boldsymbol{\xi}^k) - \eta \leq \theta^k, \ \forall k \in \mathcal{K}, \\
& \left(\mathbf{X}\boldsymbol{\xi}^k + \mathbf{x^0}, \mathbf{W}\boldsymbol{\xi}^k + \mathbf{w^0} \right) \in \mathcal{G}(\mathbf{z}, \boldsymbol{\xi}^k), \ \forall k \in \mathcal{K}, \\
& \theta^k \geq 0, \ \forall k \in \mathcal{K} \\
& \mathbf{z} \in \{0, 1\}^n.
\end{aligned}
\tag{4.12}
$$

The reformulation will effectively address the non-smoothness caused by the $(.)_+$ operation but at the cost of introducing additional variables and constraints. It is worthwhile to do that here as the latter formulation will result in an overall MILP.

5   Computational results

In this section we report the numerical experiments performed to evaluate the proposed methodologies. We have used MATLAB R2015b with CPLEX 12.6 for solving transportation problems and quadratic optimisation problems while SEDUMI 1.3 was used for solving SDPs.

We first provide numerical results on a small case study in Section 5.1 with three facilities and four customers. The purpose of this part is to illustrate the performance of various components of the SDP and SIP models such as the constraint generation algorithm and the CVaR approximation. We also use this example to demonstrate how the sample size affects the performance of the SIP model. We then report the numerical results for medium test instances in Section 5.2. Here we report the computational times for some test instances in the literature.

20

We also study the robustness of the proposed SDP and SIP solutions by varying the realised demand distributions. For convenience in recapping these methods, Table 1 provides a quick reference on their abbreviations and the key differences among them.

Abbreviation	Problem	Methods	Known information on demand
D-FLP	(2.1a)	Deterministic FLP	Actual values ($\mathbf{d}$)
S-FLP	(2.2)	Stochastic FLP	Probability distribution (P)
R-SDP	(3.16)	Robust FLP with semi-definite formulation	First and second moments ($\boldsymbol{\mu}, \mathbf{Q}$)
R-SIP	(4.4)	Robust FLP with semi-infinite formulation	
R-LDR	(4.5)	LDR approximation of R-SIP problem	First moment ($\boldsymbol{\mu}$)
R-CVaR	(4.7)	CVaR approximation of R-LDR problem	

Table 1: Abbreviations and references of methodologies used

5.1   A small case study

We study a small-scale facility location problem to illustrate the quality of solutions obtained from the proposed solution methods. We randomly generate a facility location problem with 4 demand nodes and 3 potential locations to build facilities. The transportation costs are assumed to be proportional to the distances between the customers and the facilities. Figure 1 shows the network layout of this problem.
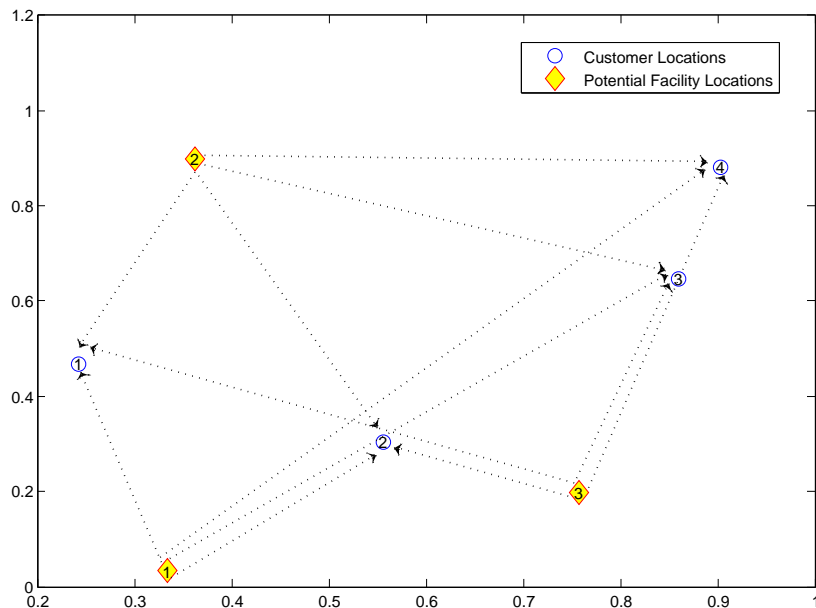

Figure 1: The network of facility location problem

The transportation cost, fixed investment cost and the capacity of each potential facility are given in Table 2.

Supply points	Demand points				Supply capacity	Fixed initial cost
	dem_1	dem_2	dem_3	dem_4		
sup_1	14	12	21	25	200	2000
sup_2	14	18	16	16	300	3200
sup_3	17	10	14	19	254	3700
External source	27	27	27	27	∞	0

Table 2: Transportation and investment costs and capacity of the facilities

We assume that customer demands are unknown prior to the construction of facilities and we are only given the first moment information of the demand with $\boldsymbol{\mu} = (150, 150, 100, 100)$. Moreover, in the SDP formulation of the problem, we are also provided with the second moment information of the uncertain demand

$$
\mathbf{Q} = \begin{pmatrix}
22669.34 & 22511.07 & 15038.4 & 15026.78 \\
22511.07 & 22551.57 & 15000.53 & 14988.58 \\
15038.4 & 15000.53 & 10045.48 & 10013.84 \\
15026.78 & 14988.58 & 10013.84 & 10031.04
\end{pmatrix}
$$

Under the assumption of available first and second moments, we formulate the problem as a robust SDP (R-SDP). Since the problem size is small, we first consider all combinations of the possible facility location decisions (2^3 possible combinations of z). For each potential solution, the full R-SDP problem (3.16) is constructed by including all of the extreme points of the dual transportation polytopes (obtained by using the signature method in Balinski and Rispoli (1993)). The full R-SDP problem is then solved for each potential solution and comparisons between the total cost of each decision are provided in Table 3.

Facility Decision	Possible FL decisions							
	sol_1	sol_2	sol_3	sol_4	sol_5	sol_6	sol_7	sol_8
z_1	0	0	0	**0**	1	1	1	1
z_2	0	0	1	**1**	0	0	1	1
z_3	0	1	0	**1**	0	1	0	1
Cost	100,132.1	56,012.3	47,851.8	**13,723.6**	64,639.2	21,349.9	13,829.1	15,507.8

Table 3: Worst expected cost associated with each possible facility location decision

Note that the very high costs of certain facility location solutions, such as with $\mathbf{z} = [0, 0, 0]$ or $\mathbf{z} = [0, 0, 1]$, are due to the high costs of using the external facility when the total (random) demand exceeds the capacity of the built facilities. It can be observed that solution number 4, i.e. $\mathbf{z} = (0, 1, 1)$ has the least cost and therefore is optimal.

In order to illustrate the performance of the proposed constraint generation algorithm for the R-SDP formulation, we first fixed the facility decisions to $\mathbf{z} = (1, 1, 1)$. In each run, the problem was initiated by randomly selecting one of the SDP constraints corresponding to one of the extreme points of the dual transportation problem. We recorded the objective value after each iteration of the constraint generation process. Figure 2 summarizes the objective value

after each iteration of the algorithm (by taking the average over 100 runs, each with a different starting point). It can be observed that the convergence of the constraint generation algorithm to the optimal value takes place after around 15 iterations.
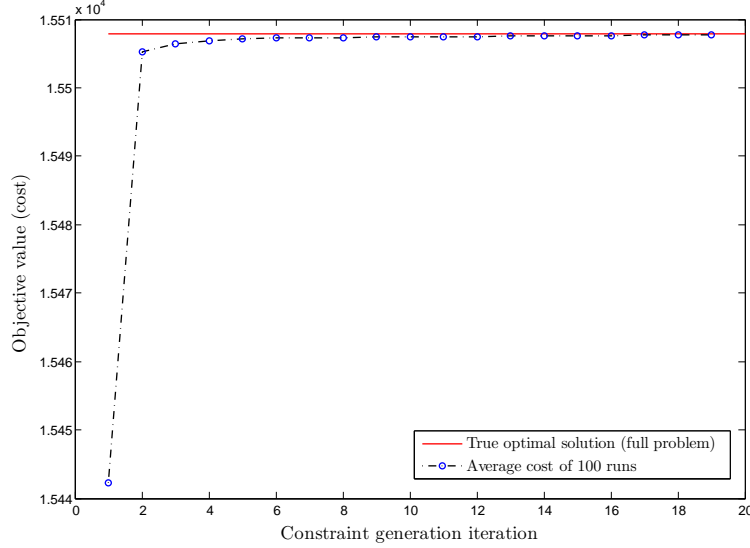

Figure 2: Convergence of the constraint generation algorithm solutions

After finding the optimal solution for a fixed facility location decision $\mathbf{z}$, the next step of the proposed solution method is to find the optimal $\mathbf{z}$ using GA. However, this is not necessary for this small instance as it has only 8 possible solutions.

5.1.1 Robust SIP formulation

Let us assume that we are only given the first moment information $\boldsymbol{\mu}$ and the second moment Q is unknown. We implement the second proposed method and formulate the FLP as a semi-infinite program. For assessing the quality of LDR approximation (R-LDR) of the "true" robust SIP (R-SIP) solution, we limit the support set of the random demand for each customer to 2000 values generated from the uniform distribution $\mathcal{U}(0, 250)$. We then solve the full R-SIP and its R-LDR approximation over this support set. As shown in Table 4, the R-LDR solution provides an upper bound approximation with 0.9% deviation from the original R-SIP solution. For comparison purposes, we also solve the deterministic version of the instance by assuming that the demand values are known and given by $\boldsymbol{\mu}$. The solution from the deterministic problem (D-FLP) is then benchmarked against the robust solution in Table 4.

Models	Optimal FLP decisions			Total cost
	z_1	z_2	z_3	
D-FLP	1	1	0	12,300.00
R-SDP	0	1	1	13,723.61
R-SIP	1	1	1	15,641.42
R-LDR	1	1	1	15,781.42

Table 4: Deterministic versus robust solutions

The deterministic solution is to install just enough capacity to meet the predicted (assumed to be known) demand values by locating facilities 1 and 2. In other words, the D-FLP solution provides no flexibility for possible variation in future demand. The robust solutions, on the other hand, offers to install the facilities with a higher total capacity at a higher total cost (i.e. constructing facilities 2 and 3 in the R-SDP case and all facilities in the R-SIP case), to hedge against the risk of not meeting the customer demand.

In the next step, we implement the proposed CVaR approximation (R-CVaR) of the R-LDR solution. Using the same support set of 2000 values, we solved the R-CVaR approximation with various β values. The results are presented in Figure 3. It can be observed that the CVaR solution approximates the R-LDR optimal solution consistently and without any error for $\beta \geq 0.7$.
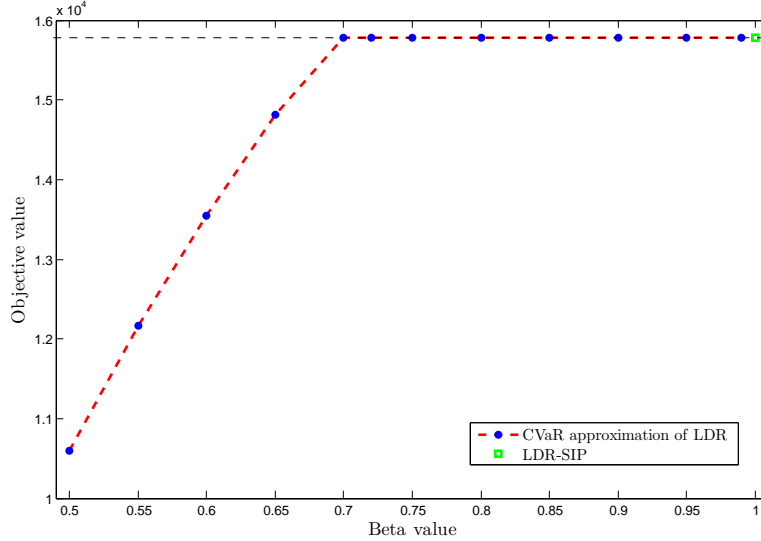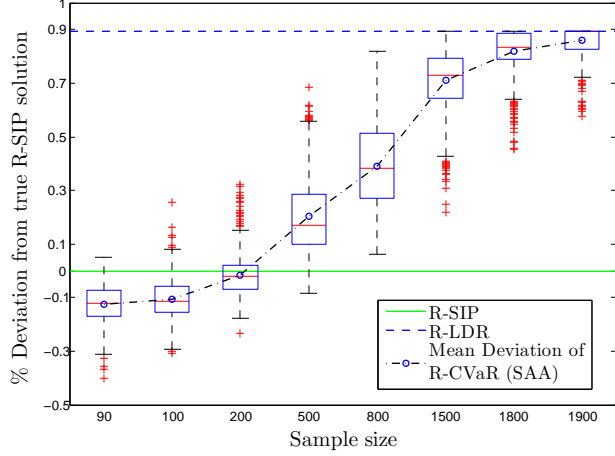

Figure 3: CVaR approximation of R-LDR problem for various β values

5.1.2 Discretization through sampling

The complexity of the R-LDR and R-CVaR schemes increases as the number of scenarios increases. As described in subsection 4.4, one way to resolve this issue is to use sample approximation (SA) for R-LDR and sample average approximation for R-CVaR. These involve drawing i.i.d samples from the underlying distribution of the uncertainty. In the case of the first instance,

we have chosen the samples from the same support set used to run the full R-SIP tests. To assess the quality of sample approximation, we solve R-SIP, R-DLR and R-CVaR using various sample sizes. For each sample size, we carried out 1000 independent runs. The β value for all of the CVaR instances was set to 0.99. The normalized deviation of the sample approximations of all problems from the true robust (full R-SIP) solution is shown in Figure 4 for various choices of the sample sizes.

(a) SAA of R-CVaR
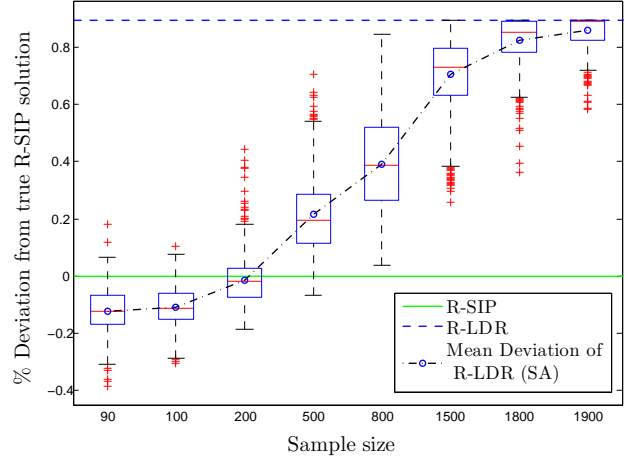
(b) SA of R-LDR
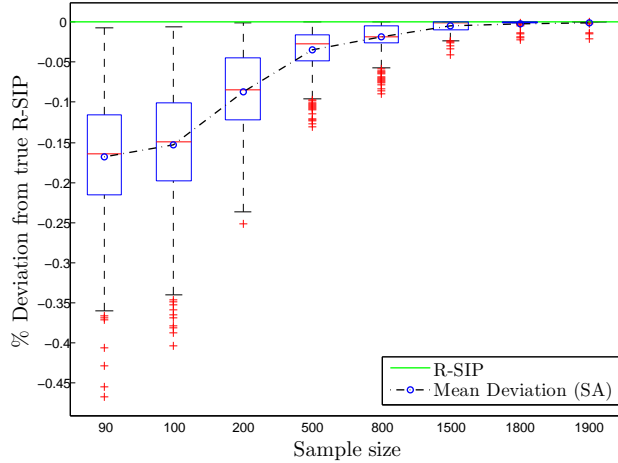
(c) SA of R-SIP

Figure 4: Normalized deviation of approximation problems from the true robust solution

It can be observed that, for each sample size, the mean deviation of R-CVaR and R-LDR solutions from the true robust solution are very similar. They range from -0.4% to 0.9% of the true solution (here 0.9% deviation means that the approximate solution is 0.9% higher than the true value of the robust solution). We can also see that the sample approximation of R-CVaR and R-LDR converge to the full R-LDR solution as the sample size increases. Furthermore, the sampling method applied to R-SIP provides a very good approximation of the true solution even

for small sample sizes.

The average computation times for 1000 runs of each sample size are shown in the graph below:
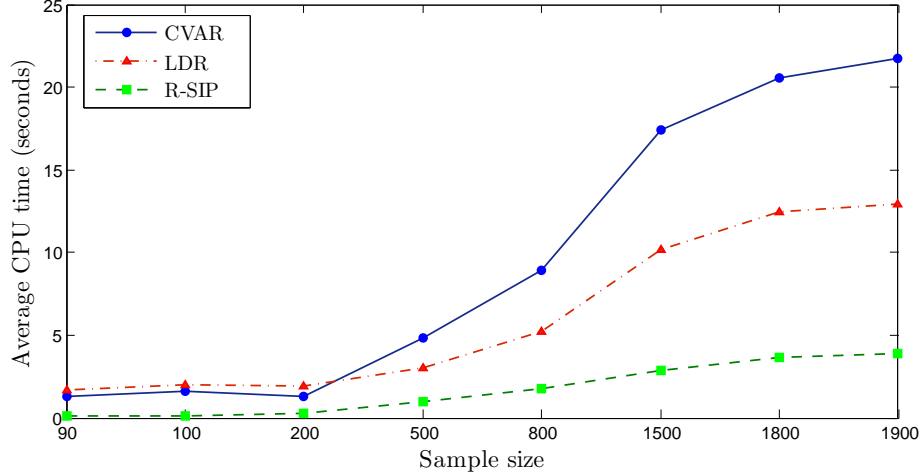

Figure 5: Average computation time

5.2 Medium size test instances

In this section, we consider a set of larger facility location problems. Test instances are selected from those existing in the literature. We have modified and used test problems presented by Díaz and Fernández (2002) for the single-source capacitated facility location problem[3]. The first 6 instances consist of networks of 10 potential facility locations and 20 customer demand points. The network size is increased twice in the last instance to observe how the computational time increases. We have used the given demand values in test instances as the first moment of the customer demand distribution. The second moment matrix for the SDP formulation was randomly generated[4].

5.2.1 SDP formulation

The test results for the medium size instances are presented in Table 5. The first two columns show the test instances from Díaz and Fernández (2002) and their corresponding network sizes. The third and fourth columns show the optimal solution **z** and the corresponding SDP objective

---

[3]Available at http://www-eio.upc.edu/~elena/sscplp/index.html. Here, we note that the facility location model in Díaz and Fernández (2002) is slightly different from Model 2 in that the authors restrict one facility to each customer, which is not appropriate in our case due to demand uncertainty. In addition, our model includes an external source which can be used if the (uncertain) demand exceeds the supply. The serving cost from the external source is set as $C = 1.2 \max_{i,j} c_{ij}$

[4]We use default Matlab function 'gallery('randcorr',m)' to generate the correlation matrix and the standard deviations is set equal to 0.04 times the first moments.

values. Columns 5-7 shows the CPU computational time while column 8 shows the optimal value of the corresponding deterministic facility location problem as defined in Model 2.

| Test | Size | Optimal robust decisions | SDP | CPU (s) | | | D-FLP |
#	(m, n)	(z)	solution	SDP	CG	Total	
p_1	$(20, 10)$	[1 1 0 1 1 0 1 1 1 0]	1,814.62	14,251	1,789	16,143	1,810.86
p_2	$(20, 10)$	[1 0 0 1 0 0 1 1 1 1]	3,903.44	11,563	1,673	13,337	3,915.73
p_3	$(20, 10)$	[0 0 0 0 0 0 0 0 0 0]	4,218.68	7,028	737	7,858	4,218.6
p_4	$(20, 10)$	[0 0 0 1 0 0 0 1 0 0]	4,746.66	7,223	830	8,136	4,746.42
p_5	$(20, 10)$	[1 1 0 0 1 1 1 0 1 1]	4,377.84	13,012	1,685	14,797	4,375.87
p_6	$(20, 10)$	[1 1 0 1 1 1 1 1 1 0]	2,221.55	13,441	1,725	15,266	2,172.59
p_7	$(30, 15)$	[0 1 0 1 1 1 0 0 1 0 1 0 0 1 0]	4,419.46	39,980	2,262	42,343	4,166.64
p_{18}	$(40, 20)$	[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0]	8,822.93	81,427	2,389	83,924	8,624.18

Table 5: Robust facility location solutions

We can see from column 7 that it took between 2-4 hours to run the instances with 20 customers and 10 facilities. This increases sharply to over 11 hours for instance p_7 and 22 hours for instance p_{18} when the number of customers and facilities are increased to $(30, 15)$ and $(40, 20)$, respectively. The total CPU times are broken down to the time to (iteratively) run the SDPs and to perform row generations (i.e. to identify violating constraints). We can see that most of the computational time is on solving the SDPs. In theory, the SDP models (3.9) and (3.16) can be solved in polynomial times for fixed numbers of facilities and customers since the problem has a linear objective function with a polynomial number of SDP constraints. While we have made a contribution in transforming the two-stage stochastic, minimax, integer optimisation problem into a single-level SDP optimisation problem, we acknowledge that there are still drawbacks in the framework. Specifically, despite recent developments in SDP programming, problems (3.9) and (3.16) do not scale well with the number of facilities and the number of customers due to the fast growth in the number of SDP constraints. For practical purposes with larger instances, if being able to find a reasonably robust solution is more crucial than optimality, our framework is still applicable and what we need to do is to set the time limits on the constraint generation algorithm and the genetic algorithms.

The numerical scheme for solving the SDP contains several fine tuning parameters. First of all, the stopping condition for the row generation algorithm is triggered either when a maximum of 100 SDP constraints are added or when the additional constraints do not lead to more than 0.1% changes on the objective value in the 5 consecutive iterations. Second, the number of iterations in the alternating procedure for finding violating constraints is set at at most 100 iterations. We implement the constraint generation and utilise the built-in Matlab GA algorithms with mostly default settings except that the GA population size is set to 30 with stopping criteria of a maximum of 10 generations. The population size is the number of solutions under consideration at any point by the GA algorithm to perform various operations in order to produce further solutions. The parameters are set to balance between the run time and the quality of the solutions. All the numerical tests are run in MATLAB R2015b utilising SeDuMi and Cplex LP solvers version 12.6.

For comparison purposes, we also include the deterministic solutions of the instances in the table above on column 8. As expected, most of the robust solutions have higher costs than deterministic facility decisions as a result of installing higher total capacities, although the lower cost of D-FLP solutions comes at the expense of non-flexibility of the facility decisions against the fluctuation in future customer demand. The only exception is on instance p_2. Here, we note that the row generation algorithm have some stopping criteria and the reported SDP solutions might not be the 'true optimal' solutions should we have enough computational power to solve the big SDPs with all the constraints included.

In order to verify the performance of GA, we compare its solutions with the exhaustive search technique where the objective values for every combination of $\mathbf{z}$ are evaluated in order to find the optimal solutions. This exhaustive search is only doable when n is small enough, i.e., $n = 10$ in the first 6 instances. We found that the solution obtained from GA matched the optimal solution in the first 5 instances while the optimality gap for instance p_6 is quite small at 1.5%[5].

5.2.2 SIP formulation

We now reconsider the problem instance p_5 and assume that the only available information on demand uncertainty is the first moment of the distribution $\boldsymbol{\mu}$. The R-SIP framework is then used to construct and solve this problem. As before, we limit the support set of the random variable $\boldsymbol{\xi}$ to 200 uniformly generated random values. For comparison purposes, we summarize the full R-SIP solution to this problem along with those from D-FLP, S-FLP and full R-SDP versions of the problem in Table 6 below.

Model	Available information	Optimal FLP decisions										Total cost	Installed capacity
		z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}		
D-FLP	$\mathbf{d}$	1	1	0	0	0	1	1	1	1	1	4417.26	458
S-FLP	P	1	1	0	1	1	1	1	0	1	0	4495.33	462
R-SDP	$(\boldsymbol{\mu}, \mathbf{Q})$	1	1	0	0	1	1	1	1	1	0	4603.65	481
R-SIP	$\boldsymbol{\mu}$	1	1	0	0	1	1	1	1	1	1	4938.95	511

Table 6: FLP solutions for various methodologies

It can be observed that R-SIP offers a more flexible solution by installing a higher level of supply capacity. This, of course, comes at the expense of a higher total expected cost. Also, as we include progressively less information on the uncertainty in our models, the solution becomes more and more robust (higher levels of supply capacity installed), which is intuitively sensible.

Having solved the full R-SIP p_5, we consider sample approximation of the R-SIP, R-LDR and R-CVaR. Various sample sizes are used for each problem and we repeat each test for 100 times. The normalized deviation from the full R-SIP solution is computed and presented in Figure 6.

---

[5]Despite the consistent performance of GA in finding the optimal solution in these instances, in theory an optimal solution cannot be guaranteed. However, in practice, the "local" solutions obtained using GA are often of high quality. For instance p_3, the robust solution coincides with the deterministic solution at $\mathbf{z} = 0$. This is because the facility building fixed costs dominate the serving costs from the external source.

(a) SAA of R-CVaR (b) SA of R-LDR
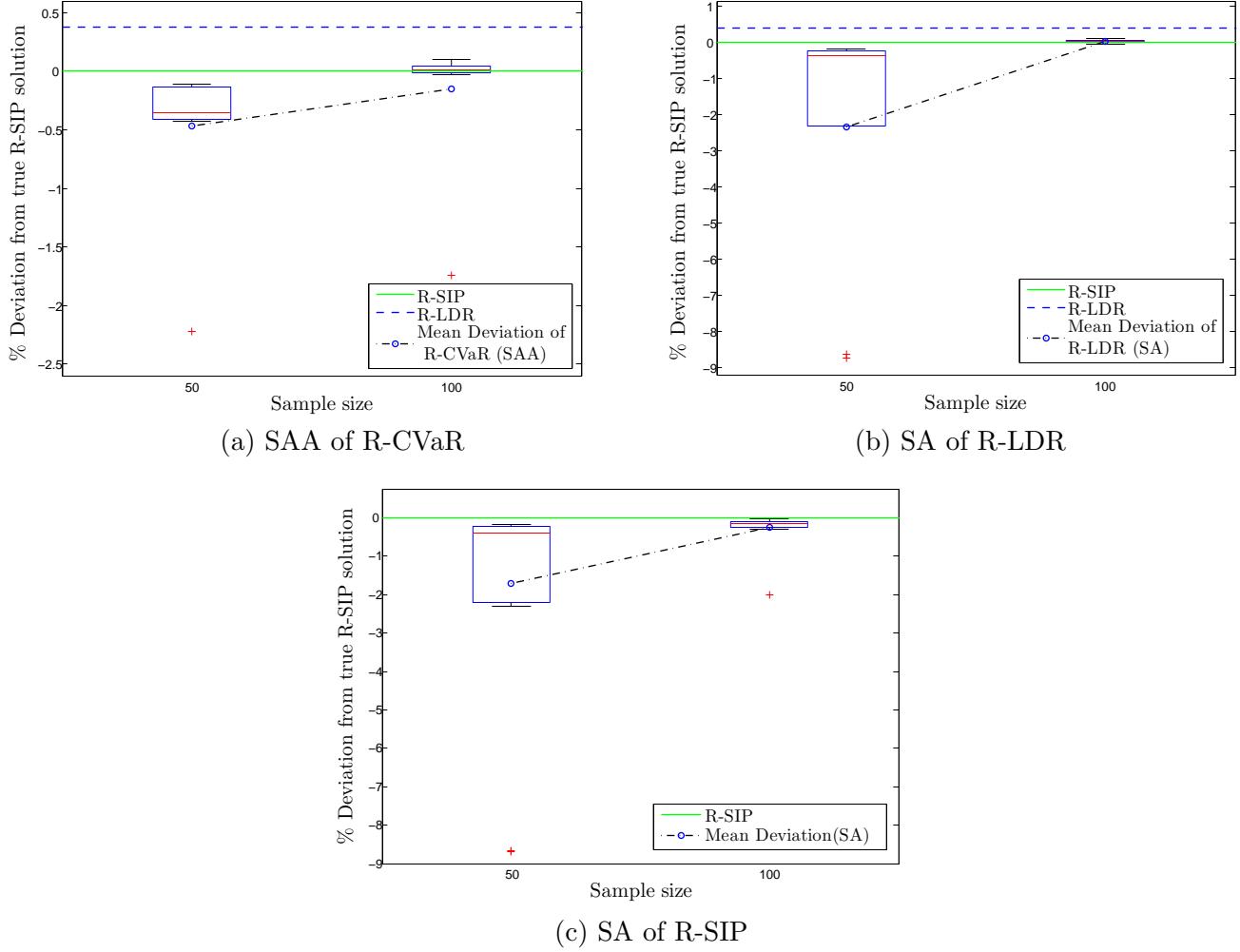

(c) SA of R-SIP

Figure 6: Deviation of approximation solutions from the true robust solution

It can be observed that, in all models, the sampling scheme results in approximation of the true robust solution (R-SIP) with a very low error (with less than 3% mean deviation). Moreover, in the case of smaller sample sizes, the CVaR model performs much better than the other two models and has a mean deviation of less than 0.5%.

6 Conclusion

In this paper we consider a capacitated facility location problem with customer demand uncertainty and propose a two-stage distributionally robust model for the problem to tackle the issue of incomplete information on the true distribution of the uncertainty. We construct the uncertainty set using the moments information associated with the distribution of the random demands. Two numerical methods are proposed based on the available moment information. Specifically, we first formulate the robust problem as a semi-definite program on the basis of the first and the second moments which is then solved by using a constraint generation algorithm.

Moreover, we formulate the robust problem as a semi-infinite program for the case that only the first moment information is given. The semi-infinite program is then solved by approximation using a linear decision rule, CVaR and Monte Carlo sampling. Finally, we carry out numerical tests for a small instance and also for some medium-sized instances taken from the literature. In each case, the distributionally robust solutions offer the flexibility in hedging against uncertainty compared to the deterministic and the stochastic solutions.

In the future, it would be interesting to study the possibility of extending the results and methodologies presented in this paper to include uncertainty in supply, multi-stage problems, and the competition in the market. Also, we would like to explore the problem structure to enhance the solution algorithms for a better performance in large-scale instances. Another open direction is to apply the proposed methodologies and numerical schemes to the practical problems with similar structure and characteristics such as uncertain supply and demand. Some examples related to the energy industry could include the wind farm site location problem and liquefied natural gas storage facility location problems.

Acknowledgements

References

E. Anderson, H. Xu, and D. Zhang. Confidence levels for cvar risk measures and minimax limits. 2014.

I. Averbakh and O. Berman. Minimax regret p-center location on a network with demand uncertainty. *Location Science*, 5(4):247–254, 1997.

I. Averbakh and O. Berman. Minmax regret median location on a network under uncertainty. *INFORMS Journal on Computing*, 12(2):104–110, 2000.

M.L. Balinski and F.J. Rispoli. Signature classes of transportation polytopes. *Mathematical programming*, 60(1-3):127–144, 1993.

M.L. Balinski and A. Russakoff. Faces of dual transportation polyhedra. In *Mathematical Programming at Oberwolfach II*, pages 1–8. Springer, 1984.

A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

D. Bertsimas and J. Sethuraman. Moment problems and semidefinite optimization. In *Handbook of semidefinite programming*, pages 469–509. Springer, 2000.

Pierre Bonami, Oktay Günlük, and Jeff Linderoth. Globally solving nonconvex quadratic programming problems with box constraints via integer programming methods. *Mathematical Programming Computation*, 10(3):333–382, 2018.

G. Calafiore and M.C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

G.C. Calafiore and M.C. Campi. The scenario approach to robust control design. *Automatic Control, IEEE Transactions on*, 51(5):742–753, 2006.

M.C. Campi and S. Garatti. A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. *Journal of Optimization Theory and Applications*, 148 (2):257–280, 2011.

Jieqiu Chen and Samuel Burer. Globally solving nonconvex quadratic programming problems via completely positive programming. *Mathematical Programming Computation*, 4(1):33–52, 2012.

E. Conde. Minmax regret location–allocation problem on a network under uncertainty. *European journal of operational research*, 179(3):1025–1039, 2007.

M.S. Daskin. What you should know about location modeling. *Naval Research Logistics (NRL)*, 55(4):283–294, 2008.

M.S. Daskin. *Network and discrete location: models, algorithms, and applications*. John Wiley & Sons, 2011.

L. Davis. Handbook of genetic algorithms. 1991.

E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

K. Derinkuyu and M. Pınar. On the s-procedure and some variants. *Mathematical Methods of Operations Research*, 64(1):55–77, 2006.

J.A. Díaz and E. Fernández. A branch-and-price algorithm for the single source capacitated plant location problem. *Journal of the Operational Research Society*, pages 728–740, 2002.

P.M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166, 2018.

Jacek Gondzio and E Alper Yildirim. Global solutions of nonconvex standard quadratic programs via mixed integer linear programming reformulations. *arXiv preprint arXiv:1810.02307*, 2018.

S.K. Goyal. Improving vam for unbalanced transportation problems. *The Journal of the Operational Research Society*, 35(12):1113–1114, 1984.

L.J. Hong, Y. Yang, and L. Zhang. Sequential convex approximations to joint chance constrained programs: A monte carlo approach. *Operations Research*, 59(3):617–630, 2011.

D. Kuhn, W. Wiesemann, and A. Georghiou. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, 130(1):177–209, 2011.

H.J. Landau. *Moments in mathematics*, volume 37. American Mathematical Soc., 1987.

M. Lim, M.S. Daskin, A. Bassamboo, and S. Chopra. A facility reliability problem: Formulation, properties, and algorithm. *Naval Research Logistics (NRL)*, 57(1):58–70, 2010.

F.V. Louveaux. Discrete stochastic location models. *Annals of Operations research*, 6(2):21–34, 1986.

D. Love and G. Bayraksan. Phi-divergence constrained ambiguous stochastic programs for data-driven optimization. *Technical report, Department of Integrated Systems Engineering, The Ohio State University, Columbus, Ohio*, 2015.

M. Lu, L. Ran, and J.M. Shen. Reliable facility location design under uncertain correlated disruptions. *Manufacturing & Service Operations Management*, 17(4):445–455, 2015.

S.H. Owen and M.S. Daskin. Strategic facility location: A review. *European Journal of Operational Research*, 111(3):423–447, 1998.

C.S. ReVelle and H.A. Eiselt. Location analysis: A synthesis and survey. *European Journal of Operational Research*, 165(1):1–19, 2005.

J. Santiváñez and H.J. Carlo. Reliable capacitated facility location problem with service levels. *EURO Journal on Transportation and Logistics*, 7(4):315–341, 2018.

H. Scarf, K.J. Arrow, and S. Karlin. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production*, 10:201–209, 1958.

A. Shapiro. On duality theory of conic linear problems. In *Semi-infinite programming*, pages 135–165. Springer, 2001.

A. Shapiro. Monte carlo sampling methods. *Handbooks in operations research and management science*, 10:353–425, 2003.

A. Shapiro and A. Nemirovski. On complexity of stochastic programming problems. In *Continuous optimization*, pages 111–146. Springer, 2005.

A. Shapiro, D. Dentcheva, and A.P. Ruszczyński. *Lectures on stochastic programming: modeling and theory*, volume 9. SIAM, 2009.

L.V. Snyder. Facility location under uncertainty: a review. *IIE Transactions*, 38(7):547–564, 2006.

L.V. Snyder and M.S. Daskin. Stochastic p-robust location problems. *IIE Transactions*, 38(11): 971–985, 2006.

H. Sun and H. Xu. Convergence analysis for distributionally robust optimization and equilibrium problems. *Mathematics of Operations Research*, 41(2):377–401, 2015.

H. Sun, H. Xu, and Y. Wang. Asymptotic analysis of sample average approximation for stochastic optimization problems with joint chance constraints via conditional value at risk and difference of convex functions. *Journal of Optimization Theory and Applications*, pages 1–28, 2012.

W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

C. Wu, D. Du, and D. Xu. An approximation algorithm for the two-stage distributionally robust facility location problem. In *Advances in Global Optimization*, pages 99–107. Springer, 2015.

Wei Xia, Juan C Vera, and Luis F Zuluaga. Globally solving nonconvex quadratic programs via linear integer programming techniques. *INFORMS Journal on Computing*, 2019.

H. Xu, Y. Liu, and H. Sun. Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. *Mathematical Programming*, 169(2):489–529, 2018.

J. Zhang, H. Xu, and L. Zhang. Quantitative stability analysis for distributionally robust optimization with moment constraints. *SIAM Journal on Optimization*, 26(3):1855–1882, 2016.

S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, 2013.