Data-Driven Distributionally Robust Capacitated Facility Location Problem

Ahmed Saif* [1] and Erick Delage†[2]

[1]Department of Industrial Engineering, Dalhousie University
[2]Department of Decision Sciences, HEC Montréal

Abstract

We study a *distributionally robust* version of the classical capacitated facility location problem with a distributional ambiguity set defined as a Wasserstein ball around an empirical distribution constructed based on a small data sample. Both single- and two-stage problems are addressed, with customer demands being the uncertain parameter. For the single-stage problem, we provide a direct reformulation into a mixed-integer program. For the two-stage problem, we develop two iterative algorithms, based on column generation, for solving the problem exactly. We also present conservative approximations based on support set relaxation for the single- and two-stage problems, an affine decision rule approximation of the two-stage problem, and a relaxation of the two-stage problem based on support set restriction. Numerical experiments on benchmark instances show that the exact solution algorithms are capable of solving large scale problems efficiently. The different approximation schemes are numerically compared and the performance guarantee of the two-stage problem's solution on out-of-sample data is analyzed.

1 Introduction

Uncertainty about the future poses a challenge for decision makers, especially when taking strategic decisions that have long-lasting implications. One of the important strategic decisions that has to be made by firms is the location of their manufacturing, service and logistical facilities. Although facility location is among the earliest and best studied problems in the literature (see, for example, the review of Hale and Moberg [12]), most early facility location models are deterministic. However, interest in uncertain location problems has grown considerably in the last decade, driven by recent advances in stochastic optimization techniques and the need for new approaches to deal rigorously with uncertain parameters such as demand and cost.

*Department of Industrial Engineering, Dalhousie University, 5269 Morris Street, Halifax, NS B3H 4R2, Canada (ahmed.saif@dal.ca).

†Department of Decision Sciences, HEC Montréal, 3000 Chemin de la Côte-Sainte-Catherine, Montréal, QC H3T 2A7, Canada (erick.delage@hec.ca).

Traditionally, there have been two main approaches to handle uncertainty in optimization problems. In *Stochastic Programming* (SP), the uncertain parameters are represented as random variables having a *known* probability distribution, whereas the cost function to be minimized is an expectation over this distribution. Despite the intuitive appeal and nice convergence properties of this approach, it often results in intractable formulations. Furthermore, the assumption of full knowledge about the underlying probability distribution, which is usually estimated from limited sample data, might lead to disappointments when the optimal solution obtained is implemented with a different sample drawn from the same population, a phenomenon known as the *optimizer's curse* [29]. On the other extreme, the *Robust Optimization* (RO) approach assumes a complete ignorance about the probability distribution of uncertain parameters. Instead, these parameters are only assumed to belong to an *uncertainty set* with some structure (*e.g.,* ellipsoid or polyhedron). Optimization is performed with respect to the *worst-case* scenario in the uncertainty set, which inevitably leads to over-conservatism and suboptimal decisions for other more-likely scenarios. Another drawback of RO is that it does not fully utilize the richness of data available to the decision maker (except, probably, for calibrating the uncertainty set).

An alternative paradigm known as *distributionally robust optimization* (DRO), that provides a unifying framework for SP and RO while aiming to overcome their deficiencies, has gained a lot of attention recently. Instead of the *white-or-black* view of the SP and RO approaches when it comes to the issue of knowing the probability distribution of uncertain parameters, DRO adopts a middle-ground approach; The probability distribution is assumed to belong to a family of distributions, referred to as the *ambiguity set*, that share certain parametric characteristics or are "close-enough" to a reference distribution. The concept itself is not new and was used in the work of Scarf [26] to tackle an ambiguity-averse newsvendor problem. However, tractable reformulations for important classes of DRO have been developed only recently based on modern results from robust optimization and statistics.

Early work on DRO has focused on ambiguity sets that satisfy certain parametric conditions, *e.g.,* limits on the distribution moments [7]. These *moment-based* approaches usually result in tractable semidefinite or conic formulations but have weak convergence properties. Emphasis of DRO research has shifted recently towards *statistical distance-based* approaches that construct ambiguity sets in the vicinity of a reference distribution. A key advantage of these approaches is that they enable observed/sampled data to be incorporated directly and effectively in the optimization problem. Since they make extensive and direct use of real data, they are usually referred to as *data-driven* DRO approaches [22]. The next section provides a review of some recent work in the area of DRO.

In this paper, we address a stochastic version of the well-known Capacitated Facility Location Problem (CFLP) when the probability distribution of demand (the uncertain parameter) is not known with certainty, but rather can be only estimated based on a finite random sample of observations. We construct a distributional ambiguity set around the *empirical* distribution formed based on the historical data such that it includes all distributions within a certain distance from the reference distribution, where distance is measured using a *Wasserstein metric*. A DRO approach is implemented to hedge against distributional ambiguity and find solutions that can provide probabilistic out-of-sample performance guarantees. Both single- and two-stage problems are considered, in which the former assumes that all decisions are made at the outset without the possibility of recourse, whereas in the latter only the location decisions must be made under distributional ambiguity but the assignment of demands to open facilities is decided after the demand becomes known. In both cases, we stipulate that the solution obtained must remain feasible for all possible realizations in the bounded support set of the demand distribution. In other words, the total capacity of open facilities must be sufficient to satisfy the highest demand within the support set. With proper selections of the support set and the norm used in the Wasserstein metric, both cases can be tractably reformulated as mixed-integer linear programs. We devise two exact solution approaches for the two-stage problem. The first approach begins by dualizing the recourse problem then uses a column-and-constraint generation algorithm to implicitly enumerate the vertices of the polyhedral feasible set of the dual recourse problem. The second approach, in contrast, uses a lifting of the support set then implements a column generation algorithm to enumerate the vertices of the lifted support set. Moreover, we propose conservative approximations of the single- and two-stage problems based on support set relaxation, another approximation that uses *affine decision rules*, and a relaxation based on support set restriction. Extensive numerical testing on benchmark was conducted to evaluate the computational efficiency of the iterative algorithms, the quality of the proposed approximations and the out-of-sample performance guarantee of the optimal solutions obtained.

The remainder of this paper is organized as follows: the next section provides brief reviews of facility location problems under uncertainty and recent advances in DRO. Section 3 presents descriptions and mathematical formulations of the single- and two-stage distributionally robust CFLP. A reformulation of the single-stage problem and solution algorithms for the two-stage problem are provided in section 4. Section 5 presents useful conservative approximations and a relaxation. Numerical experiment results are presented in section 6. Finally, Conclusions are drawn in section 7.

Notation. We use upright lower and upper case letters, respectively, for vectors and matrices. Individual elements of these vectors and matrices are denoted using *italic* versions of the same letters. For example, elements of the J-dimension vector x are denoted as x_j. Depending on the context, upper and lower case letters, respectively, might be used also to denote probability distributions (*e.g.,* F_{ξ}) and functions (*e.g.,* $g(\cdot)$). Upper case calligraphic letters are used for sets (*e.g.,* $\mathcal{X}$). We use the symbol e to denote an all-ones vector of appropriate size and e_i as the i-th column of the identity matrix. When LP duality is used, the dual/primal variables are included, between parentheses, right after their corresponding primal/dual constraints in the mathematical formulation.

2 Literature Review

2.1 Facility Location Problems under Uncertainty

Given the substantial body of literature related to this topic, the aim of this section is not to provide a comprehensive survey, but rather to shed light on the main trends in facility location problems under uncertainty and point to some representative examples. The reader is referred to the reviews of Louveaux [20], Owen and Daskin [25], Snyder [30] and the text by Correia and da Gama [5] for a detailed account of the literature.

Traditionally, uncertainty in facility location problems is represented through a finite set of scenarios, each with a known probability of occurrence, an approach that dates back to the work of Sheppard [28]. Most references have considered one of two attitudes: risk-neutral (the expected value approach) or risk-averse (the min-max approach). A representative example of the former is the stochastic programming formulation of the two-stage CFLP with stochastic demand by Laporte *et al.* [19]. On the other hand, when the probabilities of different scenarios are not available or cannot be trusted, or when the decision-maker needs to hedge against extreme scenarios, the min-max approach provides an attractive alternative. Minimizing the maximum regret, defined as the difference between the cost of a solution in a given scenario and the optimal cost of that scenario, is often considered a more suitable objective than minimizing the maximum cost over all scenarios, which is deemed too pessimistic. Serra and Marianov [27] considered robust p-median problems with both *min-max cost* and *min-max regret* objectives. In an attempt to control the conservatism of the min-max approach, Daskin *et al.* [6] used a *min-max regret* objective that considers only a subset of scenarios whose collective probability of occurrence equals at least some user-specified value α.

The last decade has witnessed significant advances in RO methods, which found their applications in facility location problems. Baron *et al.* [3] proposed tractable RO reformulations for a multi-period, revenue-maximization capacitated location-transportation-production problem. They considered both box and ellipsoidal uncertainty sets for the uncertain demand. In their single-stage model, all tactical (allocation) and operational (production) decisions are taken at time zero along with the strategic location and capacity decisions, without the possibility of recourse, which leads to overly conservative solutions. Atamtürk and Zhang [2] avoided this issue in the two-stage RO reformulation they devised for a location-transportation problem (as an example of network flow problems) with a budget uncertainty set. While the two-stage approach alleviated the over-conservatism issue of the single-stage approach, it usually leads to intractable robust counterparts. As shown in Atamtürk and Zhang [2], the problem is NP-hard even for a network flow problem on a bipartite graph. To overcome this difficulty in the case of a robust multi-period location-transportation problem, Ardestani-Jaafari and Delage [1] proposed a set of conservative approximations that reduce the flexibility of the delayed decisions and identified interesting cases for which full flexibility is unnecessary to reach optimal robust solutions.

Recognizing the merits and limitations of the expectation/SP and the min-max/RO approaches, there have been some attempts to combine them in a single framework. Using a finite set of scenarios, Snyder and Daskin [31] proposed an approach that tries to balance the long-run performance advantage of SP with the risk-aversion nature of RO in two classical facility location problems. The model minimizes the expected cost while enforcing a cap on the maximum regret over all scenarios. On the negative side, the approach inherits the intractability of both the deterministic integer problems and the SP method, added to them the complexity of the ρ-robustness constraints. As noted by the authors, for many instances finding a feasible solution, and even determining whether the instance is feasible, is difficult. More recently, Keyvanshokooh *et al.* [18] studied a closed-loop supply chain network problem where the transportation cost is represented through a finite set of scenarios, whereas demand and return quantities were assumed to belong to budgeted uncertainty sets. The two uncertainty types were combined in a two-stage robust-stochastic program.

DRO offers an attractive alternative for dealing with uncertainty that combines the advantages of RO and SP while evading their drawbacks. So far, very few references have used DRO in facility location problems. For instance, Gülpınar *et al.* [11] studied a stochastic location-inventory problem with a basestock inventory policy and a stock-out probabilistic requirement stated as a chance constraint. In one of two cases they considered, they assumed that the mean demand in each node is known while the demand distribution is ambiguous and the demand belongs to an uncertainty set. Robust counterparts were derived for symmetric (ellipsoidal), asymmetric and scenario-based uncertainty sets to reflect different risk measures. Although this work can be classified as a moment-based DRO problem, it was labeled and solved as a classical RO problem without resorting to DRO reformulation approaches. Wu *et al.* [35] studied a two-stage uncapacitated facility location problem (UFLP) under a moment-based distributional ambiguity, where the first one or two moments of the demand distribution functions are known. They showed that the linear relaxation of this problem is equivalent to that of the standard UFLP. Carlsson *et al.* [4] considered a distributionally robust version of the Euclidean travelling salesman problem (TSP) in which the distributional ambiguity set is defined, similar to our work, based on a Wasserstein metric. Their objective was to divide a territory into service districts for a fleet of vehicles when limited data is available. Our paper has a different focus on facility locations and allocation of demands to open facilities.

2.2 Data-driven Distributionally Robust Optimization

In a DRO problem, we aim to find a solution that optimizes the expected value of an objective function with respect to the *worst-case* probability distribution among a specific set of distributions. Mathematically, a minimization DRO problem can be stated as:

$$\min_{\mathrm{x}\in\mathcal{X}} \sup_{F_{\xi}\in\mathcal{D}} \mathbb{E}_{F_{\xi}}\left[h(\mathrm{x},\xi)\right]$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ is the feasible set of decision variables, $\xi \in \Xi \subseteq \mathbb{R}^m$ is a random vector that represents the uncertain parameters, and $h : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R} \cup \{-\infty, \infty\}$ is a real function.

A key ingredient of any DRO model is the ambiguity set $\mathcal{D}$. Broadly speaking, ambiguity sets can be classified into two types: *moment-based* and *statistical distance-based*. In the former, the ambiguity set encompasses all the probability distributions that satisfy certain moment constraints, typically the first and second moments. Delage and Ye [7] considered a supported moment-based DRO model where the first moment of ξ is confined in an ellipsoid set whereas the second-moment matrix lies in a positive semidefinite cone. Other moment-based models have been proposed by Wiesemann *et al.* [32] and Goh and Sim [10]. On the other hand, statistical distance-based ambiguity sets encompass probability distributions that are within a certain *distance* from a nominal distribution. Several distance metrics have been proposed in the literature for constructing ambiguity sets. A statistical distance-based ambiguity set that has drawn much attention recently and that will be utilized in the paper is that based on the *Wasserstein metric*, also referred to as the Kantorovich-Rubinstein metric [17]. The Wasserstein distance between two probability distributions F_1 and F_2 can be described as the cost of an optimal transportation plan for moving the probability mass in one so it becomes identical the other. Formally, the Wasserstein metric

$d_{\mathrm{W}} : \mathcal{M}(\Xi) \times \mathcal{M}(\Xi) \mapsto \mathbb{R}$ is defined as

$$d_{\mathrm{W}}(F_1, F_2) := \inf \left\{ \int_{\Xi^2} \|\xi_1 - \xi_2\| \Pi(\mathrm{d}\xi_1, \mathrm{d}\xi_2) \;\middle|\; \begin{array}{l} \Pi \text{ is a joint distribution of } \xi_1 \text{ and } \xi_2 \\ \text{with marginals } F_1 \text{ and } F_2 \text{ respectively} \end{array} \right\},$$

where $\| \cdot \|$ represents an arbitrary norm on $\mathbb{R}^m$ and the probability space $\mathcal{M}(\Xi)$ contains all probability distributions supported on Ξ. Given a finite set $\widehat{\Xi} := \left\{ \widehat{\xi}_1, \ldots, \widehat{\xi}_N \right\}$ of sample points, each representing a historical or predicted realization of the uncertain parameters, an empirical distribution $\widehat{F}_{\xi}^N$ can be constructed such that each discrete point in the sample set has an equal probability of $\frac{1}{N}$, *i.e.,* $\widehat{F}_{\xi}^N := \frac{1}{N} \sum_{n=1}^{N} \delta_{\widehat{\xi}_n}$, where $\delta_{\xi} : \Sigma \mapsto \{0, 1\}$, $\delta_{\widehat{\xi}_n}(\mathcal{A}) = \begin{cases} 1 & \text{if } \widehat{\xi}_n \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases}$ is a *Dirac measure* concentrating unit mass at $\widehat{\xi}_n$ and Σ is a Borel σ-algebra on Ξ. The *Wasserstein ambiguity set* $\mathcal{D}_{\varepsilon}(\widehat{F}_{\xi}^N, \Xi) := \left\{ F_{\xi} \in \mathcal{M}(\Xi) \mid d_{\mathrm{W}}(F_{\xi}, \widehat{F}_{\xi}^N) \leq \varepsilon \right\}$ includes all probability distributions supported on $\Xi \subset \mathbb{R}^m$ that are within a distance $\varepsilon \geq 0$ of the reference/emperical distribution $\widehat{F}_{\xi}^N$. Intuitively, when $\varepsilon = 0$, the ambiguity set contains only the empirical distribution and the DRO problem reduces to an SP, whereas with a very large ε, the adversary can place the entire probability mass on a vertex that maximizes the cost and the problem becomes an RO.

Despite the conceptual difficulty of DRO problems with Wasserstein ambiguity sets, tractable reformulation were developed for important special cases. Wozabal [33] exploited the property that the extreme points are discrete distributions with a fixed number of atoms to devise a finite-dimensional non-convex reformulation. Subsequently, Wozabal [34] provided a more concise closed-form expression for robustifying convex, law-invariant risk measures over a Wasserstein ambiguity set without support constraints when the objective function is linear in the uncertain parameters. Zhao and Guan [37] reformulated a two-stage, data driven DRO problem as a two-stage SP with a finite support and devised a Benders approach to solve it efficiently. Gao and Kleywegt [9] showed that the data-driven DRO can be approximated by a robust program to any accuracy, and that the robust program approximation becomes exact when the objective function is concave in ξ. Mohajerin-Esfahani and Kuhn [23] showed that when the uncertainty set is convex and closed and the objective function is a point-wise minimum of concave functions, the problem can be reformulated as a finite convex program. Hanasusanto and Kuhn [13] showed that a two-stage Wasserstein-based DRO problem can be reformulated as a copositive program if the problem has complete recourse and l_2-norm is used in the Wasserstein metric definition. They also provide a linear programming reformulation when the distribution's support is unconstrained and an l_1-norm is used to define the Wasserstein ball. Recently, Luo and Mehrotra [21] proposed a cutting-surface method to solve the reformulated problem for the general nonlinear model.

3 Single- and two-stage distributionally robust capacitated facility location problems

The capacitated facility location problem (CFLP) is a classical problem that has been extensively studied in the literature. Nonetheless, for completeness, we provide the following description of the problem: We are given a set of demand points indexed by $i = 1, \ldots, I$, and a set of potential facility locations indexed by $j = 1, \ldots, J$. ξ_i denotes the demand originating from point i. The entire demand has to be served by facilities opened in a subset of potential locations. Each facility has a set-up cost f_j and a capacity v_j that determines the maximum demand quantity it can serve. There is a unit shipping cost c_{ij} between demand point i and facility location j. We aim to determine the number and locations of facilities to open and how to allocate demand to them in order to minimize the total cost, which includes set-up and shipping costs. Due to the capacitated nature of the problem, single assignment of demand points to facilities might not be optimal (or even feasible). To formulate the problem, we use two types of variables: Binary location variables $\mathrm{x} \in \{0,1\}^J$, where x_j takes value one if a facility is opened in potential location j and zero otherwise, and non-negative continuous allocation variables $\mathrm{y} \in \mathbb{R}_+^{I\times J}$, where y_{ij} represents the percentage of point i's demand served by the facility opened in location j. With that, the CFLP can be stated as

$$\min_{\mathrm{x}\in\{0,1\}^J,\, \mathrm{y}\in\mathbb{R}_+^{I\times J}} \quad \sum_{j=1}^{J} f_j x_j + \sum_{i=1}^{I}\sum_{j=1}^{J} \xi_i c_{ij} y_{ij} \tag{1a}$$

$$\text{s.t.} \quad \sum_{j=1}^{J} y_{ij} = 1 \qquad i = 1, \ldots, I \tag{1b}$$

$$\sum_{i=1}^{I} \xi_i y_{ij} \leq v_j x_j \qquad j = 1, \ldots, J. \tag{1c}$$

Assumption 1. *In the Wasserstein distributional ambiguity set $\mathcal{D}_\varepsilon(\widehat{F}_\upxi^N, \Xi)$, (i) the support set is a bounded polyhedron defined as $\Xi := \left\{\upxi \in \mathbb{R}^I \mid \mathrm{C}\upxi \leq \mathrm{d}\right\}$, for some $\mathrm{C} \in \mathbb{R}^{L\times I}$ and $\mathrm{d} \in \mathbb{R}^L$; and (ii) the norm used in the Wasserstein metric definition is an l_1-norm.*

Hence, with a *risk-averse* decision-maker who aims to avoid future disappointments and desires a probabilistic guarantee on the out-of-sample performance, we can formulate the DRO problem as

$$\min_{\mathrm{x}\in\{0,1\}^J,\, \mathrm{y}\in\mathbb{R}_+^{I\times J}} \quad \sum_{j=1}^{J} f_j x_j + \sup_{F_\upxi \in \mathcal{D}_\varepsilon(\widehat{F}_\upxi^N, \Xi)} \mathbb{E}_{F_\upxi}\left[\sum_{i=1}^{I}\sum_{j=1}^{J} \xi_i c_{ij} y_{ij}\right] \tag{2a}$$

$$\text{s.t.} \quad \sum_{j=1}^{J} y_{ij} = 1 \qquad i = 1, \ldots, I \tag{2b}$$

$$\sup_{\upxi\in\Xi} \sum_{i=1}^{I} \xi_i y_{ij} \leq v_j x_j \qquad j = 1, \ldots, J. \tag{2c}$$

The robust constraint (2c) replaces the deterministic constraint (1c) to ensure that the solution remains feasible for all possible demand realizations. In this formulation, we assume that all decisions are taken *before* uncertainty is revealed. Therefore, we refer to this problem as the *single-stage distributionally robust capacitated facility location problem (1-DR-CFLP).*

This formulation is suitable when the allocation of demands to open facilities must be decided at the outset without a possibility of *recourse*, due to, for example, contractual commitments with shipping companies or customers. Furthermore, the single stage model gives rise to a simple rule that can be implemented in a decentralized manner as each facility only needs to know the demand at the retailers that it serves. However, when only the strategic location decision needs to be made under uncertainty whereas the assignment of demands to open facilities can be made after the demands become known, the decision maker can achieve a less-conservative solution by solving the *two-stage distributionally robust capacitated facility location problem (2-DR-CFLP)*. The first-stage problem, which has the *here-and-now* location decision variables, can be stated as

$$\min_{\mathrm{x}\in\{0,1\}^J} \sum_{j=1}^{J} f_j x_j + \sup_{F_{\xi}\in\mathcal{D}_{\varepsilon}(\widehat{F}_{\xi}^N,\Xi)} \mathbb{E}_{F_{\xi}}\left[g(\mathrm{x},\xi)\right], \tag{3}$$

where $g(\mathrm{x},\xi)$ is the *recourse function*, evaluated after uncertainty is revealed by solving the second-stage problem

$$g(\mathrm{x},\xi) := \min_{\mathrm{z}\in\mathbb{R}_+^{I\times J}} \quad \sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij} z_{ij} \tag{4a}$$

$$\text{s.t.} \quad \sum_{j=1}^{J} z_{ij} \geq \xi_i \qquad i=1,\ldots,I \tag{4b}$$

$$\sum_{i=1}^{I} z_{ij} \leq v_j x_j \qquad j=1,\ldots,J. \tag{4c}$$

Note that for the two-stage problem, we utilized a well-known alternative formulation of the CFLP that uses the allocation variable $z_{ij} = \xi_i y_{ij}$, which denotes the demand of point i served by the facility opened in location j, instead of y_{ij} (See, for example, [24]). With this formulation, the second-stage problem has a *fixed recourse* and the uncertain parameters affect the right-hand side only. It also highlights the convexity of $g(\mathrm{x},\xi)$ with respect to ξ.

We also need to ensure that the recourse problem remains feasible for all feasible first-stage decisions, *i.e.,* has a *relatively complete recourse*. We enforce this property by adding the following valid inequality to the first-stage problem:

$$\sup_{\xi\in\Xi} \mathrm{e}^{\intercal}\xi \leq \mathrm{v}^{\intercal}\mathrm{x},$$

which, under Assumption 1(i), can be tractably reformulated through LP duality with multipliers w as

$$\begin{aligned} \inf_{\mathrm{w}\in\mathbb{R}_+^L} \mathrm{d}^{\intercal}\mathrm{w} \quad &\leq \quad \mathrm{v}^{\intercal}\mathrm{x} \\ \mathrm{C}^{\intercal}\mathrm{w} \quad &= \quad \mathrm{e}. \qquad\qquad (\xi) \end{aligned}$$

4 Reformulations and Exact Solution Approaches

In this section, we show how to tractably reformulate the single- and two-stage distributionally-robust CFLP, and provide two exact solution approaches for the latter.

4.1 Reformulation of the 1-DR-CFLP

For a given feasible solution $(\bar{\mathrm{x}}, \bar{\mathrm{y}})$, the inner maximization in the 1-DR-CFLP's objective function (2a) can be tractably reformulated directly, using the result from Mohajerin Esfahani and Kuhn [23, Corollary 5.1], as

$$\min_{\lambda,\, \gamma \in \mathbb{R}_+^{N\times L}} \quad \varepsilon\lambda + \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{i=1}^{I}\sum_{j=1}^{J}\widehat{\xi}_{in} c_{ij}\bar{y}_{ij} + \sum_{l=1}^{L}\left(d_l - \sum_{i=1}^{I} C_{li}\widehat{\xi}_{in}\right)\gamma_{ln}\right) \tag{5a}$$

$$\text{s.t.} \quad \left\| \sum_{i=1}^{I}\left(\sum_{l=1}^{L} C_{li}\gamma_{ln} - \sum_{j=1}^{J} c_{ij}\bar{y}_{ij}\right)\mathrm{e}_i \right\|_* \leq \lambda \qquad n = 1,\ldots,N. \tag{5b}$$

Under Assumption 1(ii), the dual norm $\|.\|_*$ is an l_∞-norm and constraint (5b) reduces to

$$\left| \sum_{l=1}^{L} C_{li}\gamma_{ln} - \sum_{j=1}^{J} c_{ij}\bar{y}_{ij} \right| \leq \lambda \qquad i = 1,\ldots,I,\ n = 1,\ldots,N.$$

Moreover, the robust constraint (2c) can be tractably reformulated through LP duality as

$$\min_{\mathrm{u} \in \mathbb{R}_+^{J\times L}} \quad \sum_{l=1}^{L} d_l u_{jl} \leq v_j \bar{x}_j \qquad j = 1,\ldots,J$$

$$\text{s.t.} \quad \sum_{l=1}^{L} C_{li} u_{jl} = \bar{y}_{ij} \qquad i = 1,\ldots,I,\ j = 1,\ldots,J.$$

Hence, the 1-DR-CFLP is equivalent to

$$
\begin{aligned}
\min_{\mathrm{x},\mathrm{y},\lambda,\gamma,\mathrm{u}} \quad & \sum_{j=1}^{J} f_j x_j + \varepsilon\lambda + \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{i=1}^{I}\sum_{j=1}^{J}\widehat{\xi}_{in} c_{ij} y_{ij} + \sum_{l=1}^{L}\left(d_l - \sum_{i=1}^{I} C_{li}\widehat{\xi}_{in}\right)\gamma_{ln}\right) \\
\text{s.t.} \quad & \sum_{l=1}^{L} C_{li}\gamma_{ln} - \sum_{j=1}^{J} c_{ij} y_{ij} \le \lambda && i=1,\dots,I,\ n=1,\dots,N \\
& \sum_{j=1}^{J} c_{ij} y_{ij} - \sum_{l=1}^{L} C_{li}\gamma_{ln} \le \lambda && i=1,\dots,I,\ n=1,\dots,N \\
& \sum_{j=1}^{J} y_{ij} = 1 && i=1,\dots,I \\
& \sum_{l=1}^{L} d_l u_{jl} \le v_j x_j && j=1,\dots,J \\
& \sum_{l=1}^{L} C_{li} u_{jl} = y_{ij} && i=1,\dots,I,\ j=1,\dots,J \\
& \mathrm{x} \in \{0,1\}^J,\ \mathrm{y} \in \mathbb{R}_+^{I\times J}, \gamma \in \mathbb{R}_+^{N\times L}, \mathrm{u} \in \mathbb{R}_+^{J} \times L.
\end{aligned}
$$

This is a mixed-integer linear program (MILP) with only J binary variables that can be easily solved using commercial solvers.

4.2 Exact reformulations and solution approaches for the 2-DR-CFLP

This section presents two exact solution approaches for the two-stage distributionally-robust CFLP. In the first approach, the problem is reformulated after dualizing the recourse problem into a large-scale mixed-integer linear program, before a column-and-constraint generation algorithm is used to solve it. In the second approach, a lifting of the support set is utilized to reformulate the problem and a column generation algorithm is developed for the reformulated problem. In both approaches, we use results from [23] to deal with the DRO problem.

4.2.1 A column-and-constraint generation algorithm with a dualized recourse problem

We begin the reformulation of the 2-DR-CFLP by dualizing constraints (4b) and (4c) with multipliers $\nu \in \mathbb{R}_+^I$ and $\mu \in \mathbb{R}_+^J$, respectively. However, to have a bounded dual recourse problem, we first need to ensure feasibility of the primal recourse problem by writing it as

$$g(\mathrm{x}, \xi) := \min_{\mathrm{z} \in \mathbb{R}_+^{I \times J}, \theta \geq 0} \quad \sum_{i=1}^{I} \sum_{j=1}^{J} c_{ij} z_{ij} + \bar{c}\theta \tag{6a}$$

$$\text{s.t.} \quad \sum_{j=1}^{J} z_{ij} \geq \xi_i \qquad i = 1, \ldots, I \qquad (\nu_i) \tag{6b}$$

$$\sum_{i=1}^{I} z_{ij} \leq v_j x_j + \theta \qquad j = 1, \ldots, J, \qquad (\mu_j) \tag{6c}$$

where $\bar{c} \geq \max_{i,j}(c_{ij})$ is a scalar large enough to ensure that $\theta^* = 0$ unless the selection of the first-stage decision x makes the recourse problem infeasible. Next, through LP duality we get

$$g(\mathrm{x}, \xi) = \max_{\nu \in \mathbb{R}_+^I, \mu \in \mathbb{R}_+^J} \quad \sum_{i=1}^{I} \xi_i \nu_i - \sum_{j=1}^{J} v_j x_j \mu_j \tag{7a}$$

$$\text{s.t.} \quad \nu_i \leq \mu_j + c_{ij} \qquad i = 1, \ldots, I, \; j = 1, \ldots, J \quad (z_{ij}) \tag{7b}$$

$$\sum_{j=1}^{J} \mu_j \leq \bar{c}\,. \qquad (\theta) \tag{7c}$$

Note that the polyhedral feasible set of the dual problem (7) is bounded, as it can be contained in the hypercube $\left\{ (\nu, \mu) \;:\; \nu_i \in [0, \max_j c_{ij} + \bar{c}], \; i = 1, \ldots, I, \; \mu_j \in [0, \bar{c}], \; j = 1, \ldots, J \right\}$, and that it depends on neither x nor ξ. Hence, we can solve the dual problem by enumeration over its vertices. Let $\mathcal{K} := \{k\}_{k=1}^{K}$ be the index set of vertices of the dual feasible set

$$\mathcal{V} := \{(\nu, \mu) : \; 0 \leq \nu_i \leq \mu_j + c_{ij}, \, 0 \leq \mu_j \leq \bar{c}, \, i = 1, \ldots, I, \, j = 1, \ldots, J\}\,.$$

Therefore, the dual problem can be written as $g(\mathrm{x}, \xi) = \max_{k \in \mathcal{K}} \left(\sum_{i=1}^{I} \xi_i \nu_i^k - \sum_{j=1}^{J} v_j x_j \mu_j^k \right)$. Using the result from Mohajerin Esfahani and Kuhn [23, Corollary 5.1], the 2-DR-CFLP can be tractably reformulated as

$$\min_{\mathrm{x,w,\lambda,s,\gamma}} \quad \sum_{j=1}^{J} f_j x_j + \lambda\varepsilon + \frac{1}{N}\sum_{n=1}^{N} s_n \tag{8a}$$

$$\text{s.t.} \quad -\sum_{j=1}^{J} v_j x_j \mu_j^k + \sum_{i=1}^{I} \widehat{\xi}_{in}\nu_i^k + \sum_{l=1}^{L}\left(d_l - \sum_{i=1}^{I} C_{li}\widehat{\xi}_{in}\right)\gamma_{lnk} \le s_n \qquad n = 1,\dots,N,\ k \in \mathcal{K} \tag{8b}$$

$$\|\mathrm{C}^\intercal \boldsymbol{\gamma}_{nk} - \boldsymbol{\nu}^k\|_* \le \lambda \qquad n = 1,\dots,N,\ k \in \mathcal{K} \tag{8c}$$

$$\mathrm{d}^\intercal \mathrm{w} \le \mathrm{v}^\intercal \mathrm{x} \tag{8d}$$

$$\mathrm{C}^\intercal \mathrm{w} = \mathrm{e} \tag{8e}$$

$$\mathrm{x} \in \{0,1\}^J, \mathrm{w} \in \mathbb{R}_+^L, \mathrm{s} \in \mathbb{R}^N, \boldsymbol{\gamma} \in \mathbb{R}_+^{N\times L\times K}. \tag{8f}$$

Given the exponential size of $\mathcal{K}$, we begin with small subsets $\mathcal{K}'_n \subset \mathcal{K}$ for constrains indexed by $n = 1,\dots,N$ and employ a column-and-constraint generation algorithm to generate and add new vertices to $\mathcal{K}'_n$ iteratively. We first write problem (8) as

$$\begin{aligned}
\min_{\mathrm{x,w,\lambda,s}} \quad & \sum_{j\in J} f_j x_j + \lambda\varepsilon + \frac{1}{N}\sum_{n=1}^{N} s_n \\
\text{s.t.} \quad & h_n(\mathrm{x}, \lambda, \boldsymbol{\mu}^k, \boldsymbol{\nu}^k) \le s_n \qquad n = 1,...,N,\ k \in \mathcal{K} \\
& \mathrm{d}^\intercal \mathrm{w} \le \mathrm{v}^\intercal \mathrm{x} \\
& \mathrm{C}^\intercal \mathrm{w} = \mathrm{e} \\
& \mathrm{x} \in \{0,1\}^J, \mathrm{w} \in \mathbb{R}_+^L, \mathrm{s} \in \mathbb{R}^N,
\end{aligned}$$

where, for a given $k \in \mathcal{K}$,

$$h_n(\mathrm{x}, \lambda, \boldsymbol{\mu}, \boldsymbol{\nu}) := \min_{\boldsymbol{\gamma}_n \in \mathbb{R}_+^L} \quad -\sum_{j=1}^{J} v_j x_j \mu_j + \sum_{i=1}^{I} \widehat{\xi}_{in}\nu_i + \left(\mathrm{d} - \mathrm{C}\widehat{\boldsymbol{\xi}}_n\right)^\intercal \boldsymbol{\gamma}_n \tag{9a}$$

$$\text{s.t.} \quad \|\mathrm{C}^\intercal \boldsymbol{\gamma}_n - \boldsymbol{\nu}\|_* \le \lambda. \tag{9b}$$

Under Assumption 1(ii), (9b) reduces to

$$\begin{aligned}
-\mathrm{C}^\intercal \boldsymbol{\gamma}_n + \boldsymbol{\nu} &\le \lambda \qquad (\boldsymbol{\psi}^+ \in \mathbb{R}^I) \\
\mathrm{C}^\intercal \boldsymbol{\gamma}_n - \boldsymbol{\nu} &\le \lambda. \qquad (\boldsymbol{\psi}^- \in \mathbb{R}^I)
\end{aligned}$$

For a given n, we are looking for the index of a new vertex $(\boldsymbol{\mu}^{k'}, \boldsymbol{\nu}^{k'})$ to be added to $\mathcal{K}'_n$ such that $h_n(\mathrm{x}, \lambda, \boldsymbol{\mu}^{k'}, \boldsymbol{\nu}^{k'}) > s_n$. If no such vertex exists for any $n = 1,\dots,N$, we conclude that x is optimal with respect to problem (8). Therefore, for given $\bar{\lambda}$ and $\bar{\mathrm{x}}$, we need to solve $\max_{k\in\mathcal{K}} h_n(\bar{\mathrm{x}}, \bar{\lambda}, \boldsymbol{\mu}^k, \boldsymbol{\nu}^k)$. Yet, since $h_n(\mathrm{x}, \lambda, \boldsymbol{\mu}, \boldsymbol{\nu})$ is jointly convex with respect to $(\boldsymbol{\mu}, \boldsymbol{\nu})$, this is equivalent to solving

$$\max_{\nu\in\mathbb{R}^I,\mu\in\mathbb{R}^J} \quad h_n(\bar{x},\bar{\lambda},\mu,\nu) \tag{10a}$$

$$\text{s.t.} \quad 0 \le \nu_i \le \mu_j + c_{ij} \qquad i=1,\dots,I,\ \forall j=1,\dots,J \tag{10b}$$

$$0 \le \sum_{j=1}^{J} \mu_j \le \bar{c}. \tag{10c}$$

Similar to what has been done in [36], by applying the KKT optimality conditions to the inner minimization (9), the separation subproblem (10) can be reformulated as the mixed-integer linear program

$$\bar{s}_n(\bar{x},\bar{\lambda}) = \max_{\mu,\nu,\gamma,\psi^+,\psi^-,B^0,B^+,B^-} \ -\sum_{j=1}^{J} v_j x_j \mu_j + \sum_{i=1}^{I} \widehat{\xi}_{in}\nu_i + (d - C\widehat{\xi}_n)^\intercal \gamma \tag{11a}$$

$$\text{s.t.} \quad \mathbf{0} \le d - C(\widehat{\xi}_n - \psi^+ + \psi^-) \le MB^0 \tag{11b}$$

$$\mathbf{0} \le \gamma \le M(e - B^0) \tag{11c}$$

$$\mathbf{0} \le \bar{\lambda} + C^\intercal\gamma - \nu \le MB^+ \tag{11d}$$

$$\mathbf{0} \le \psi^+ \le M(e - B^+) \tag{11e}$$

$$\mathbf{0} \le \bar{\lambda} - C^\intercal\gamma + \nu \le MB^- \tag{11f}$$

$$\mathbf{0} \le \psi^- \le M(e - B^-) \tag{11g}$$

$$0 \le \nu_i \le \mu_j + c_{ij} \qquad i=1,\dots,I,\ j=1,\dots,J \tag{11h}$$

$$0 \le \sum_{j\in J} \mu_j \le \bar{c} \tag{11i}$$

$$\mu \in \mathbb{R}^J,\ \nu \in \mathbb{R}^I,\ \gamma \in \mathbb{R}^L,\ \psi^+,\psi^- \in \mathbb{R}^I \tag{11j}$$

$$B^0 \in \{0,1\}^L,\ B^+,B^- \in \{0,1\}^I. \tag{11k}$$

Note that the binary variables B^0, B^+ and B^- are used to linearize the complementarity constraints

$$\left(d_l - \sum_{i=1}^{I} C_{li}(\widehat{\xi}_{in} - \psi_i^+ + \psi_i^-)\right)\gamma_l = 0 \qquad l=1,\dots,L$$

$$\left(\bar{\lambda} + \sum_{l=1}^{L} C_{li}\gamma_l - \nu_i\right)\psi_i^+ = 0 \qquad i=1,\dots,I$$

$$\left(\bar{\lambda} - \sum_{l=1}^{L} C_{li}\gamma_l + \nu_i\right)\psi_i^- = 0 \qquad i=1,\dots,I,$$

respectively, resulting from applying the KKT optimality conditions.

The optimal value of the relaxed master problem (problem (8) with the subsets $\mathcal{K}'_n \subset \mathcal{K}, n = 1,\dots,N$) provides a lower bound LB for the optimal value of (8). Furthermore, by solving (11) for any feasible $\bar{x}$ and $\bar{\lambda}$, we obtain an upper bound. The algorithm is initiated with any feasible $\bar{x}$ and $\bar{\lambda}$. In every iteration, separation subproblems (11) are solved for every $n = 1, \dots, N$ to find new vertices and to update the upper bound as $UB \leftarrow \min\left(UB, \sum_{j\in J} f_j\bar{x}_j + \varepsilon\bar{\lambda} + \frac{1}{N}\sum_{n=1}^{N}\bar{s}_n\right)$. The relaxed master problem is then solved with all the vertices added so far to update the values of $\bar{\mathrm{x}}$ and $\bar{\lambda}$, which are used in the next iteration, and to obtain a lower bound. If the gap between the best upper bound obtained so far and the current lower bound becomes small enough, the algorithm terminates and the incumbent x is declared optimum. A pseudocode of the column-and-constraint generation algorithm is shown in Algorithm 1.

Input: $\left\{\widehat{\xi}_n,\ n = 1, \dots, N\right\} \subset \Xi$, $\bar{\mathrm{x}}$ feasible to 1-DR-CFLP, $\varepsilon \geq 0$, $\epsilon \geq 0$
Output: ϵ-optimal solution to 2-DR-CFLP
Initialize $\lambda \geq 0$, $\mathcal{K}'_n \longleftarrow \emptyset$, $\forall n = 1, \dots, N$, $LB \longleftarrow -\infty$, $UB \longleftarrow \infty$
while $UB - LB > \epsilon$ **do**
  $\forall n \in N$, solve the subproblem (11) to get a new vertex $(\mu^{k'}, \nu^{k'})$ and $\bar{s}_n(\bar{\mathrm{x}}, \bar{\lambda})$
  $\mathcal{K}'_n \longleftarrow \mathcal{K}'_n \cup \{k'\}$ for all $n = 1, \dots, N$
  $UB \longleftarrow \min\left(UB, \sum_{j=1}^{J} f_j\bar{x}_j + \bar{\lambda}\varepsilon + \frac{1}{N}\sum_{n=1}^{N}\bar{s}_n\right)$
  Solve the relaxed master problem (8) with the index subsets $\mathcal{K}'_n,\ n = 1, \dots, N$ to obtain a new $\bar{\mathrm{x}}$ and $\bar{\lambda}$ and update LB
end

Algorithm 1: A column-and-constraint generation algorithm with a dualized recourse problem

4.2.2 A column generation algorithm with a support set lifting

An alternative reformulation of the 2-DR-CFLP can be obtained starting from a step in the proof of Theorem 4.2 in [23] which shows that for a given $\bar{\mathrm{x}}$, the term $\sup_{F_\xi \in \mathcal{D}_\varepsilon(\widehat{F}_\xi^N, \Xi)} \mathbb{E}_{F_\xi}\left[g(\bar{\mathrm{x}}, \xi)\right]$ is equivalent to

$$\min_{\lambda\in\mathbb{R}_+, \mathrm{s}\in\mathbb{R}^N} \quad \lambda\varepsilon + \frac{1}{N}\sum_{n=1}^{N} s_n \tag{12a}$$

$$\text{s.t.} \quad \max_{\xi\in\Xi}\left(g(\bar{\mathrm{x}}, \xi) - \lambda\|\xi - \widehat{\xi}_n\|\right) \leq s_n \qquad n = 1, \dots, N. \tag{12b}$$

In our case, since the recourse function $g(\mathrm{x}, \xi)$ and the norm $\|\xi - \widehat{\xi}_n\|$ are both convex functions in ξ, the objective in (12b) is a *Difference of Convex* (DC) function [16]. Hence, the inner maximization is a non-convex optimization problem and LP strong duality can not be utilized. Instead, we introduce the lifted space $\Xi' := \Xi'_1 \times \cdots \times \Xi'_N$ to represent uncertainty, in which

$$\Xi'_n := \left\{(\xi, \zeta) \in \mathbb{R}^{I+1} \;\middle|\; \begin{array}{l} \xi \in \Xi \\ \|\xi - \widehat{\xi}_n\| \leq \zeta \leq \bar{\zeta}_n \end{array}\right\},$$

where $\bar{\zeta}_n = \sup_{\xi\in\Xi}\|\xi - \widehat{\xi}_n\|$ is chosen such that $\zeta \leq \bar{\zeta}_n$ makes Ξ'_n bounded while preserving the exactness of the reformulation. Therefore, the robust constraint (12b) can be stated as

$$\max_{(\xi,\zeta)\in\Xi'_n} (g(\bar{x},\xi) - \lambda\zeta) \leq s_n \qquad n = 1,\dots,N.$$

It is easy to show that when Assumption 1 holds, the lifted support set Ξ' is a polyhedron. Since the inner problem in (12b) is a convex maximization over a bounded polyhedral set, its maximum is attained at one of the vertices of Ξ'_n. Hence, we replace the maximization over Ξ'_n with an equivalent maximization over the set $\left\{(\xi^{k_n},\zeta^{k_n})\right\}_{k_n\in\mathcal{K}_n}$ of vertices for each $n = 1.\dots,N$, where $\mathcal{K}_n$ is the index set for the vertices of Ξ'_n. Thus, the 2-DR-CFLP is equivalent to

$$\begin{aligned}
\min_{x\in\{0,1\}^J, w\in\mathbb{R}^L_+, \lambda\in\mathbb{R}_+, s\in\mathbb{R}^N} \quad & \sum_{j=1}^{J} f_j x_j + \lambda\varepsilon + \frac{1}{N}\sum_{n=1}^{N} s_n && \text{(13a)}\\
\text{s.t.} \quad & g(x,\xi^{k_n}) - \lambda\zeta^{k_n} \leq s_n \qquad \forall k_n \in \mathcal{K}_n,\ n = 1,\dots,N && \text{(13b)}\\
& d^\intercal w \leq v^\intercal x && \text{(13c)}\\
& C^\intercal w = e. && \text{(13d)}
\end{aligned}$$

Obviously, it is not practical to consider the entire index sets $\mathcal{K}_n$, $n = 1,\dots,N$ at the outset. Instead, we employ a column generation algorithm to successively generate and incorporate new elements of $\mathcal{K}$ as needed [8]. The algorithm is initialized with feasible $\bar{x}$ and $\bar{\lambda} \geq 0$, and with empty index sets $\mathcal{K}'_n = \emptyset$, $n = 1,\dots,N$. In every iteration, and for $n = 1,\dots,N$, we solve the separation subproblem

$$\bar{s}_n = \sup_{(\xi,\zeta)\in\Xi'_n} g(x,\xi) - \lambda\zeta \tag{14}$$

to generate a new vertex $(\xi^{k'_n},\zeta^{k'_n})$. Index of the newly generated vertex is added to $\mathcal{K}'_n$, *i.e.*, $\mathcal{K}'_n \leftarrow \mathcal{K}'_n \cup \{k'\}$, and the upper bound is updated as $UB \leftarrow \min\left(UB, \sum_{j\in J} f_j\bar{x}_j + \varepsilon\bar{\lambda} + \frac{1}{N}\sum_{n=1}^{N}\bar{s}_n\right)$. We then solve the relaxed master problem (problem (13) with the subsets $\mathcal{K}'_n \subset \mathcal{K}_n$, $n = 1,\dots N$) to update the values of $\bar{x}$ and $\bar{\lambda}$ and to set its optimal value as an updated lower bound LB. The algorithm iterates between solving the separation subproblem and solving the relaxed master problem until the gap between the upper and lower bounds becomes sufficiently small. We note that the column generation algorithm described in this section is quite similar to the column-and-constraint generation algorithm presented in Section 4.2.1. Thus, a pseudocode similar to the one depicted in Algorithm 1 can be written for the column generation algorithm by replacing the step in which subproblem (11) is solved to get a new vertex $(\mu^{k'},\nu^{k'})$ by one in which subproblem (14) is solved to get a new vertex $(\xi^{k'_n},\zeta^{k'_n})$, and by solving the relaxed master problem (13) instead of (8).

Finite convergence of the algorithm is guaranteed given that the polyhedral support set Ξ'_n has a finite number of vertices. What remains now is to show how to solve the subproblem (14) with a given $(\bar{x},\bar{\lambda})$ and for a specific n. By substituting the definition of the recourse function from (4), the subproblem expands to

$$
\begin{array}{llll}
\sup\limits_{\xi,\zeta,\delta} \min\limits_{z\in\mathbb{R}_+^{I\times J}} & \sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J} c_{ij}z_{ij} - \bar{\lambda}\zeta & & \\
\text{s.t.} & \sum\limits_{j=1}^{J} z_{ij} = \xi_i & i=1,\ldots,I & (\nu_i) \\
& \sum\limits_{i=1}^{I} z_{ij} \le v_j\bar{x}_j & j=1,\ldots,J & (\mu_j) \\
& \sum\limits_{i=1}^{I} C_{li}\xi_i \le d_l & l=1,\ldots,L & \\
& \xi_i - \widehat{\xi}_{in} \le \delta_i & i=1,\ldots,I & \\
& \widehat{\xi}_{in} - \xi_i \le \delta_i & i=1,\ldots,I & \\
& \sum\limits_{i=1}^{I} \delta_i \le \zeta & & \\
& 0 \le \zeta \le \bar{\zeta}_n. & &
\end{array}
$$

By dualizing the inner minimization problem we get the bilinear program

$$
\begin{array}{llll}
\sup\limits_{\xi,\zeta,\delta} \max\limits_{\nu\in\mathbb{R}_+^{I},\mu\in\mathbb{R}^{I}} & \sum\limits_{i=1}^{I} \nu_i\xi_i - \sum\limits_{j=1}^{J} \mu_j v_j\bar{x}_j - \bar{\lambda}\zeta & & \\
\text{s.t.} & \nu_i - \mu_j \le c_{ij} & i=1,\ldots,I,\ j=1,\ldots,J & \\
& \sum\limits_{i=1}^{I} C_{li}\xi_i \le d_l & l=1,\ldots,L & (\alpha_l) \\
& \xi_i - \widehat{\xi}_{in} \le \delta_i & i=1,\ldots,I & (\psi_i^+) \\
& \widehat{\xi}_{in} - \xi_i \le \delta_i & i=1,\ldots,I & (\psi_i^-) \\
& \sum\limits_{i=1}^{I} \delta_i \le \zeta & & (\beta) \\
& 0 \le \zeta \le \bar{\zeta}_n. & & (\gamma)
\end{array}
$$

We then swap the two maximizations and replace the inner optimization with its equivalent KKT conditions while using the dual objective to get the MILP

$$
\begin{array}{lll}
\sup\limits_{\substack{\xi,\zeta,\delta,\nu,\mu,\alpha,\beta,\gamma,\psi^+,\psi^- \\ B_1,B_2,B_3,B_4,B_5,B_6,B_7,B_8}} & d^\intercal\alpha + \widehat{\xi}_n^\intercal(\psi^+ - \psi^-) + \bar{\zeta}_n\gamma - \sum_{j=1}^{J}\mu_j v_j \bar{x}_j & \\
\text{s.t.} & \nu_i - \mu_j \le c_{ij} & i = 1,\ldots,I,\ j = 1,\ldots,J \\
& \mathbf{0} \le d - C\xi \le M B_1 & \\
& \mathbf{0} \le \alpha \le M(e - B_1) & \\
& \mathbf{0} \le \delta - \xi + \widehat{\xi}_n \le M B_2 & \\
& \mathbf{0} \le \psi^+ \le M(e - B_2) & \\
& \mathbf{0} \le \delta + \xi - \widehat{\xi}_n \le M B_3 & \\
& \mathbf{0} \le \psi^- \le M(e - B_3) & \\
& 0 \le \zeta - e^\intercal\delta \le M B_4 & \\
& 0 \le \beta \le M(1 - B_4) & \\
& 0 \le \bar{\zeta}_n - \zeta \le M B_5 & \\
& 0 \le \gamma \le M(1 - B_5) & \\
& \mathbf{0} \le C^\intercal\alpha + \psi^+ - \psi^- - \nu \le M B_6 & \\
& \mathbf{0} \le \xi \le M(e - B_6) & \\
& \mathbf{0} \le \beta - \psi^+ - \psi^- \le M B_7 & \\
& \mathbf{0} \le \delta \le M(e - B_7) & \\
& 0 \le \gamma - \beta + \bar{\lambda} \le M B_8 & \\
& 0 \le \zeta \le M(1 - B_8) & \\
& \xi,\delta,\nu,\psi^+,\psi^- \in \mathbb{R}^I,\ \mu \in \mathbb{R}^J,\ \alpha \in \mathbb{R}^L,\ \beta,\gamma,\zeta \in \mathbb{R} & \\
& B_1 \in \{0,1\}^L, B_2, B_3, B_6, B_7 \in \{0,1\}^I, B_4, B_5, B_8 \in \{0,1\}. &
\end{array}
$$

5 Conservative Approximations and Relaxations

In this section, we show how upper and lower bounds on the optimal values of the single- and two-stage DR-CFLP could be obtained by solving conservative approximations and relaxations, respectively.

5.1 Conservative approximations of 1-DR-CFLP and 2-DR-CFLP using support relaxation

The exact algorithms described in section 4.2 require that N subproblems, each having a large number of binary variables, be solved in every iteration. Although decomposability into multiple subproblems makes the algorithm amenable to parallelization, the problem might become computationally challenging as the sample size increases. Likewise, one might find that the reformulation proposed in section 4.1 for 1-DR-CFLP becomes computationally demanding as N is increased. For this reason, it can become interesting to conservatively approximate these two models. In this section, we propose conservative approximations for both models, inspired by the work of Mohajerin Esfahani and Kuhn [23], that rely on relaxing the support constraint in $\mathcal{D}_\varepsilon(\widehat{F}_\xi^N, \Xi)$. These relaxations will use the following result.

Lemma 1. *(Theorem 6.3 in [23]) Let $\ell_\mathrm{x}(\xi)$ be a proper, convex, and lower semicontinuous function. Then, for any $\varepsilon \geq 0$, we have that*

$$\sup_{F_\xi \in \mathcal{D}_\varepsilon(\widehat{F}_\xi^N, \mathbb{R}^I)} \mathbb{E}_{F_\xi}[\ell_\mathrm{x}(\xi)] = \mathbb{E}_{\widehat{F}_\xi^N}[\ell_\mathrm{x}(\xi)] + \varepsilon Lip(\mathrm{x})$$

where $Lip(\mathrm{x}) := \sup_{\rho:\ell_\mathrm{x}^*(\rho)<\infty} \|\rho\|_*$ is referred as the Lipschitz modulus of $\ell_\mathrm{x}(\cdot)$, and $\ell_\mathrm{x}^*(\rho)$ is the convex conjugate of $\ell_\mathrm{x}(\xi)$.

Starting with 1-DR-CFLP, we replace $\mathcal{D}_\varepsilon(\widehat{F}_\xi^N, \Xi)$ with $\mathcal{D}_\varepsilon(\widehat{F}_\xi^N, \mathbb{R}^I)$ in problem (2). When applying Lemma 1, we consider the function $\ell_\mathrm{y}(\xi) := \sum_{i=1}^I \sum_{j=1}^J \xi_i c_{ij} y_{ij}$ with convex conjugate $\ell_\mathrm{y}^*(\rho) := \mathbf{1}\{\rho_i = \sum_{j=1}^J c_{ij} y_{ij}\}$, *i.e.*, the indicator function that returns zero if $\rho_i = \sum_{j=1}^J c_{ij} y_{ij}$ and infinity otherwise. This implies that $\mathrm{Lip}(\mathrm{y}) = \| \sum_{i=1}^I \sum_{j=1}^J \mathrm{e}_i c_{ij} y_{ij} \|_*$. Hence, the conservative approximation for 1-DR-CFLP that is based on relaxation of the support information reduces to:

$$\min_{\mathrm{x,y,u}} \quad \sum_{j=1}^J f_j x_j + \sum_{i=1}^I \sum_{j=1}^J \bar{\xi}_i c_{ij} y_{ij} + \varepsilon \| \sum_{i=1}^I \sum_{j=1}^J \mathrm{e}_i c_{ij} y_{ij} \|_* \tag{15a}$$

$$\text{s.t.} \quad \sum_{j=1}^J y_{ij} = 1 \qquad i = 1, \ldots, I \tag{15b}$$

$$\sum_{l=1}^L d_l u_{jl} \leq v_j x_j \qquad j = 1, \ldots, J \tag{15c}$$

$$\sum_{l=1}^L C_{li} u_{jl} = y_{ij} \qquad i = 1, \ldots, I, j = 1, \ldots, J \tag{15d}$$

$$\mathrm{x} \in \{0,1\}^J, \mathrm{y} \in \mathbb{R}_+^{I \times J}, \mathrm{u} \in \mathbb{R}^{J \times L}, \tag{15e}$$

where $\bar{\xi} := (1/N) \sum_{n=1}^N \hat{\xi}^n$ is the empirical mean of ξ. This problem can be reformulated as a linear program when l_1-norm (following Assumption 1(ii)) is used, in which case $\| \sum_{i=1}^I \sum_{j=1}^J \mathrm{e}_i c_{ij} y_{ij} \|_* = \max_i \sum_j c_{ij} y_{ij}$.

Following with the 2-DR-CFLP, the bound that we get by replacing $\mathcal{D}_\varepsilon(\widehat{F}_\xi^N, \Xi)$ with $\mathcal{D}_\varepsilon(\widehat{F}_\xi^N, \mathbb{R}^I)$ in problem (3) will actually depend on the definition that is employed for $g(\mathrm{x}, \xi)$. It is clear for instance, that the conservative approximation becomes useless when using the definition in (4) since then for any $\mathrm{x} \in \mathbb{R}_+^J$ and positive ε, there exists an $F \in \mathcal{D}_\varepsilon(\widehat{F}_\xi^N, \mathbb{R}^I)$ with positive mass on some ξ for which the problem in $g(\mathrm{x}, \xi)$ becomes infeasible. We therefore need to employ an equivalent definition of $g(\mathrm{x}, \xi)$ that is feasible for any $\xi \in \mathbb{R}^I$ and makes $\ell_\mathrm{x}(\xi) := g(\mathrm{x}, \xi)$ a proper, convex, and lower semicontinuous function. With this in mind, we choose to employ the definition in (6).

Based on Lemma 1, one can establish that

$$\sup_{F_\xi \in \mathcal{D}_\varepsilon(\widehat{F}_\xi^N, \mathbb{R}^I)} \mathbb{E}_{F_\xi}\left[g(\mathrm{x}, \xi)\right] = \mathbb{E}_{\widehat{F}_\xi^N}\left[g(\mathrm{x}, \xi)\right] + \varepsilon \mathrm{Lip}(\mathrm{x}),$$

where $\mathrm{Lip}(\mathrm{x}) := \sup_{\rho : \ell_{\mathrm{x}}^*(\rho) < \infty} \|\rho\|_*$. Using the supremum representation of $g(\mathrm{x}, \xi)$ we get that the condition $\ell_{\mathrm{x}}^*(\rho) < \infty$ can be reformulated as:

$$\begin{aligned}
\infty > \ell_{\mathrm{x}}^*(\rho) &= \sup_{\xi} \rho^\intercal \xi - \ell_{\mathrm{x}}(\xi) \\
&= \sup_{\xi} \inf_{\nu, \mu : \nu_i \leq \mu_j + c_{ij}, \sum_j \mu_j \leq \bar{c}, \nu \geq 0, \mu \geq 0} \rho^\intercal \xi - \sum_{i=1}^{I} \xi_i \nu_i + \sum_{j=1}^{J} v_j x_j \mu_j \\
&= \inf_{\nu, \mu : \nu_i \leq \mu_j + c_{ij}, \sum_j \mu_j \leq \bar{c}, \nu \geq 0, \mu \geq 0} \sup_{\xi} \rho^\intercal \xi - \sum_{i=1}^{I} \xi_i \nu_i + \sum_{j=1}^{J} v_j x_j \mu_j \\
&= \inf_{\mu : \rho_i \leq \mu_j + c_{ij}, \sum_j \mu_j \leq \bar{c}, \rho \geq 0, \mu \geq 0} \sum_{j=1}^{J} v_j x_j \mu_j,
\end{aligned}$$

where the order of the $\sup_\xi$ and $\inf_{\nu,\mu}$ operations can be reversed according to Sion's minimax theorem since the function is bilinear in ξ and (ν, μ) and since the feasible space for (ν, μ) is convex and bounded.

We therefore conclude that the condition is therefore satisfied if and only if there exists a feasible μ. Hence, we get that

$$\begin{aligned}
\mathrm{Lip}(\mathrm{x}) = \bar{\mathrm{L}} := \sup_{\rho, \mu} \quad & \|\rho\|_* \\
\text{s.t.} \quad & \rho_i \leq \mu_j + c_{ij} \qquad i = 1, \ldots, I, j = 1, \ldots, J \\
& \sum_{j=1}^{J} \mu_j \leq \bar{c} \\
& \mu \geq \mathbf{0}, \rho \geq \mathbf{0}.
\end{aligned}$$

Note that $\mathrm{Lip}(\mathrm{x})$ is actually independent of x and can easily be evaluated when Assumption 1(ii) holds, in which case $\|\rho\|_* = \|\rho\|_\infty = \max_i \rho_i$. Overall, the conservative approximation for 2-DR-CFLP that is based on relaxation of the support information reduces to

$$\min_{\mathrm{x,w,z}} \quad \sum_{j=1}^{J} f_j x_j + \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij} z_{ij}^n + \varepsilon \bar{\mathrm{L}} \tag{16a}$$

$$\text{s.t.} \quad \sum_{j=1}^{J} z_{ij}^n \geq \widehat{\xi}_i^n \qquad i = 1, \ldots, I, n = 1, \ldots, N \tag{16b}$$

$$\sum_{i=1}^{I} z_{ij}^n \leq v_j x_j \qquad j = 1, \ldots, J, \forall n = 1, \ldots, N \tag{16c}$$

$$z_{ij}^n \geq 0 \qquad i = 1, \ldots, I, j = 1, \ldots, J, n = 1, \ldots, N \tag{16d}$$

$$\mathrm{d}^\intercal \mathrm{w} \leq \mathrm{v}^\intercal \mathrm{x} \tag{16e}$$

$$\mathrm{C}^\intercal \mathrm{w} = \mathrm{e} \tag{16f}$$

$$\mathrm{x} \in \{0,1\}^J, \mathrm{w} \in \mathbb{R}_+^L, \mathrm{z} \in \mathbb{R}^{I \times J \times N}. \tag{16g}$$

5.2 Conservative approximations of 2-DR-CFLP using affine decision rules

Another alternative for obtaining a conservative approximation in the case of 2-DR-CFLP is to use affine decision rules (ADR). With ADR, an assumption is made that the recourse variables depend affinely on the realized value of the uncertain parameter. To use this approximation, we start from the reformulated problem (12) and apply ADR on each constraint (12b) individually. In particular, we have that $\max_{\xi \in \Xi}\left(g(\mathrm{x}, \xi) - \lambda \|\xi - \widehat{\xi}_n\|\right)$ is bounded from above by

$$\min_{p_{ij}^n \in \mathbb{R},\, \mathrm{q}_{ij}^{n\,\intercal} \in \mathbb{R}^I} \max_{\xi \in \Xi} \left(\sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}(p_{ij}^n + \mathrm{q}_{ij}^n \xi) - \lambda \|\xi - \widehat{\xi}_n\| \right),$$

subject to the feasibility constraints on p and q outlined below. With that, the Affinely-adjustable 2-DR-CFLP with a Wasserstein ambiguity set can be reformulated as

$$\min_{\mathrm{x,p,q},\lambda,\mathrm{s}} \sum_{j=1}^{J} f_j x_j + \varepsilon\lambda + \frac{1}{N}\sum_{n=1}^{N} s_n \tag{17a}$$

$$\text{s.t.} \quad \sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}(p_{ij}^n + \mathrm{q}_{ij}^n \xi) - \lambda\|\xi - \widehat{\xi}_n\| \le s_n \qquad n = 1,\dots,N,\ \forall \xi \in \Xi \tag{17b}$$

$$\sum_{j=1}^{J} \left(p_{ij}^n + \mathrm{q}_{ij}^n \xi\right) \ge \xi_i \qquad i = 1,\dots,I,\ n = 1,\dots,N,\ \forall \xi \in \Xi \tag{17c}$$

$$\sum_{i=1}^{I} \left(p_{ij}^n + \mathrm{q}_{ij}^n \xi\right) \le v_j x_j \qquad j = 1,\dots,J,\ n = 1,\dots,N,\ \forall \xi \in \Xi \tag{17d}$$

$$p_{ij}^n + \mathrm{q}_{ij}^n \xi \ge 0 \qquad i = 1,\dots,I,\ j = 1,\dots,J,\ n = 1,\dots,N,\ \forall \xi \in \Xi \tag{17e}$$

$$\mathrm{x} \in \{0,1\}, \mathrm{p} \in \mathbb{R}^{I\times J\times N}, \mathrm{Q} \in \mathbb{R}^{I\times J\times N\times I}, \ \mathrm{s} \in \mathbb{R}^N.$$

Under assumption 1, constraints (17b) can be reformulated using LP-duality as

$$\begin{aligned}
&\sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}p_{ij}^n + \sum_{i\in I}\widehat{\xi}_{in}\left(\psi_{in}^+ - \psi_{in}^-\right) + \sum_{l\in L} d_l\gamma_{ln} \le s_n && n = 1,\dots,N\\
&\psi_{\omega n}^+ - \psi_{\omega n}^- + \sum_{l\in L} C_{l\omega}\gamma_{ln} = \sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij} q_{\omega ij}^n && \omega = 1,\dots,I,\ n = 1,\dots,N\\
&\psi_{in}^+ + \psi_{in}^- \le \lambda && i = 1,\dots,I,\ n = 1,\dots,N\\
&\psi^+, \psi^- \in \mathbb{R}_+^{I\times N},\ \gamma \in \mathbb{R}_+^{L\times N},
\end{aligned}$$

with the dual variables ψ^+, ψ^- and γ added to the minimization. Note that the index ω is used for the elements of q_{ij}, to differentiate it from the subscript i. To finalize the reformulation, we reformulate the robust constraints (17c)-(17e) using LP duality. Under Assumption 1, the robust constraint (17c) indexed by (i,n) is dualized using multiplier $\mathrm{v}_{in} \in \mathbb{R}_+^L$ as

$$\sum_{j=1}^{J} p_{ij}^n - \sum_{l=1}^{L} d_l v_{iln} \ge 0 \tag{18a}$$

$$\sum_{l=1}^{L} C_{l\omega} v_{iln} + \sum_{j=1}^{J} q_{\omega ij}^n \ge 0 \qquad \omega = 1,\dots,I,\ \omega \ne i \tag{18b}$$

$$\sum_{l=1}^{L} C_{l\omega} v_{\omega ln} + \sum_{j=1}^{J} q_{\omega ij}^n \ge 1 \qquad \omega = i. \tag{18c}$$

Likewise, the robust constraint (17d) indexed by (j,n) is dualized using multipliers $\mathrm{u}_{jn} \in \mathbb{R}^L_+$ as

$$\sum_{i=1}^{I} p_{ij}^n + \sum_{l=1}^{L} d_l u_{jln} \le v_j x_j \tag{19a}$$

$$\sum_{l=1}^{L} C_{l\omega} u_{jln} \ge \sum_{i=1}^{I} q_{\omega ij}^n \qquad \omega = 1, \ldots, I. \tag{19b}$$

And finally, the robust constraint (17e) indexed by (i,j,n) is dualized using multiplier $\mathrm{o}_{ijn} \in \mathbb{R}^L_+$ as

$$p_{ij}^n - \sum_{l=1}^{L} d_l o_{ijln} \ge 0 \tag{20a}$$

$$\sum_{l=1}^{L} C_{l\omega} o_{ijln} + q_{\omega ij}^n \ge 0 \qquad \omega = 1, \ldots, I. \tag{20b}$$

Constraints (17c), (17d) and (17e) are replaced with constraint sets (18), (19) and (20), respectively, and the new auxiliary variables v_{in}, u_{jn} and o_{ijn} are included in the optimization.

5.3 Lower bound for 2-DR-CFLP

Given that solving 2-DR-CFLP can potentially be numerically challenging, one might be interested in confirming the quality of the conservative approximations proposed in this section by comparing their optimal value to a lower bound. In particular, a lower bound for 2-DR-CFLP can be obtained by restricting the worst-case distribution to be supported on the same support as the empirical distribution, *i.e.,* $\widehat{\Xi}$. This gives rise to the following robust linear program:

$$\min_{\mathrm{x}\in\{0,1\}^J, \mathrm{w}\in\mathbb{R}^L_+, z\in\mathbb{R}^{I\times J\times N}_+} \quad \sum_{j=1}^{J} f_j x_j + \sup_{\mathrm{r}\in\mathcal{R}} \sum_{n=1}^{N} r_n \left(\sum_{i=1}^{I} \sum_{j=1}^{J} c_{ij} z_{ij}^n \right) \tag{21a}$$

$$\text{s.t.} \quad \sum_{j=1}^{J} z_{ij}^n \ge \widehat{\xi}_{in} \qquad i = 1, \ldots, I,\ n = 1, \ldots, N \tag{21b}$$

$$\sum_{i=1}^{I} z_{ij}^n \le v_i x_i \qquad j = 1, \ldots, J,\ n = 1, \ldots, N \tag{21c}$$

$$\mathrm{d}^\intercal \mathrm{w} \le \mathrm{v}^\intercal \mathrm{x} \tag{21d}$$

$$\mathrm{C}^\intercal \mathrm{w} = \mathrm{e}, \tag{21e}$$

where

$$
\mathcal{R} := \left\{ \mathrm{r} \in \mathbb{R}^N \;\middle|\; \exists\, \boldsymbol{\pi} \in \mathbb{R}_+^{N\times N}, \quad \begin{array}{ll} r_m = \frac{1}{N}\sum_{n=1}^{N} \pi_{nm} & m = 1,\dots,N \\ \sum_{m=1}^{N} \pi_{nm} = 1 & n = 1,\dots,N \\ \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{N} \|\widehat{\xi}_n - \widehat{\xi}_m\| \pi_{nm} \le \varepsilon & \end{array} \right\}.
$$

The inner maximization problem in (21a) can be written as

$$
\begin{aligned}
\sup_{\pi \in \mathbb{R}_+^{N\times N}} \quad & \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{N} \pi_{nm} \left(\sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij} z_{ij}^n \right) && && \text{(22a)} \\
\text{s.t.} \quad & \sum_{m=1}^{N} \pi_{nm} = 1 && n = 1,\dots,N \quad (\gamma_n) && \text{(22b)} \\
& \sum_{n=1}^{N}\sum_{m=1}^{N} \pi_{nm} \|\widehat{\xi}_n - \widehat{\xi}_m\| \le N\varepsilon. && (\eta) && \text{(22c)}
\end{aligned}
$$

By dualizing (22) with multipliers $\boldsymbol{\gamma}$ and η and reintegrating in (21), we obtain the linear program

$$
\begin{aligned}
\min_{\mathrm{x,w,z},\boldsymbol{\gamma},\eta} \quad & \sum_{j=1}^{J} f_j x_j + \sum_{n=1}^{N} \gamma_n + N\varepsilon\eta && && \text{(23a)} \\
\text{s.t.} \quad & \text{(21b)} - \text{(21e)} && && \\
& \frac{1}{N}\left(\sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij} z_{ij}^n \right) \le \gamma_m + \|\widehat{\xi}_n - \widehat{\xi}_m\| \eta && n = 1,\dots,N, m = 1,\dots,N && \\
& && && \text{(23b)} \\
& \mathrm{x} \in \{0,1\}^J, \mathrm{w} \in \mathbb{R}_+^L, \mathrm{z} \in \mathbb{R}_+^{I\times J\times N}, \boldsymbol{\gamma} \in \mathbb{R}^N, \eta \ge 0. && && \text{(23c)}
\end{aligned}
$$

The optimal value of (23) provides a lower bound for 2-DR-CFLP.

6 Numerical Experiments

We conducted a series of numerical tests to validate and compare the different DRO mathematical formulations proposed and to evaluate the performance of the exact solution approaches developed for the 2-DR-CFLP problem. All solution approaches were coded in Matlab R2017a and solved using Gurobi 9.0.1 on a PC with an Intel Core i7-7700 @ 3.6 GHz processor and 16 GB of RAM.

We tested on benchmark instances from [15], commonly known as the *Holmberg* test instances. Eight instances are used, and their sizes ($J \times I$) are as follows: $p02$ and $p04$: 10×50, $p14$ and $p16$: 20×50, $p30$ and $p32$: 30×150, $p56$ and $p58$: 30×200. Each pair of equal sized instances differ only in the setup cost of facilities, where these costs are 300 and 700, respectively, for the instances in the first three pairs, whereas the last pair has setup costs of 500 and 1500, respectively. Herein, we refer to these benchmark instances as *deterministic instances*. Three sample sizes of $N = 12$, 24, and 48, representing one, two and four years of monthly demands, respectively, were used in the experiments. For each combination of deterministic instance and sample size, we generated 10 *stochastic instances*, each with its own support set size and demand distribution. The support set Ξ in every stochastic instance is constructed as follows:

$$\Xi := \left\{ \xi \in \mathbb{R}^I \,:\, \xi_i^{nom}(1 - MaxDev_i) \leq \xi_i \leq \xi_i^{nom}(1 + MaxDev_i),\, \forall i \in I \right\},$$

where ξ^{nom} is the nominal/expected demand extracted from the deterministic instance and $MaxDev_i$, $i = 1, \ldots, I$ is the maximum deviation, drawn uniformly from the interval $[0.5, 1]$. Furthermore, the demand at each node is assumed to be independent and follows a $Beta(a_i, b_i)$ probability distribution supported on the interval $[\xi_i^{nom}(1 - MaxDev_i), \xi_i^{nom}(1 + MaxDev_i)]$ with the distribution parameters a_i and b_i drawn uniformly from $[0, 2]$. Using a *Beta* distribution with random parameters generates a wide range of distributional patterns, all supported on Ξ. For instance, when $a = b = 1$, the *Beta* distribution reduces to the uniform distribution. Once the demand distributions for stochastic instances are determined, we create three *DRO instances* (one for each sample size) from each stochastic instance by randomly drawing N demand realizations $\widehat{\xi}_n$ (or simply *realizations*), also referred to as *in-sample* data, to construct the empirical distribution $\widehat{F}_\xi^N$. We also randomly draw $n_{OS} = 100N$ realizations according to the same Beta distributions and refer to them as *out-of-sample* data, to be used in section 6.3. Finally, for each DRO instance, we tested for different values of ε, the Wasserstein ball radius, ranging between 0 and 400, with increments of 40. Hence, a total of 2640 tests were conducted.

6.1 Computational efficiency of exact methods

We first tested the computational performance of the two iterative algorithms proposed to solve the 2-DR-CFLP. Figures 1 and 2, respectively, show the average computational times of the column-and-constraint-generation algorithm (section 4.2.1) and the column generation with a support lifting algorithm (section 4.2.2) when applied to solve the DRO instances to optimality. It can be seen that both algorithms were able to solve DRO instances generated based on all deterministic instances, except $p32$, in less than one hour per instance on average. The most difficult DRO instance was solved in slightly less than 6000 seconds by both algorithms. In fact, both algorithms exhibited quite similar computational times and took similar number of iterations to solve each instance, which implies that they are equivalent in some sense. However, investigating, and possibly proving, this equivalence is beyond the scope of this paper. It has been found that CPU times increased by a factor of 2.07 on average when the sample size N is doubled (from 12 to 24 or from 24 to 48), which suggests that the computational time might be approximately linear in sample size. Overall, the results obtained in this round of experiments confirm that the proposed algorithms can solve problems of realistic sizes.

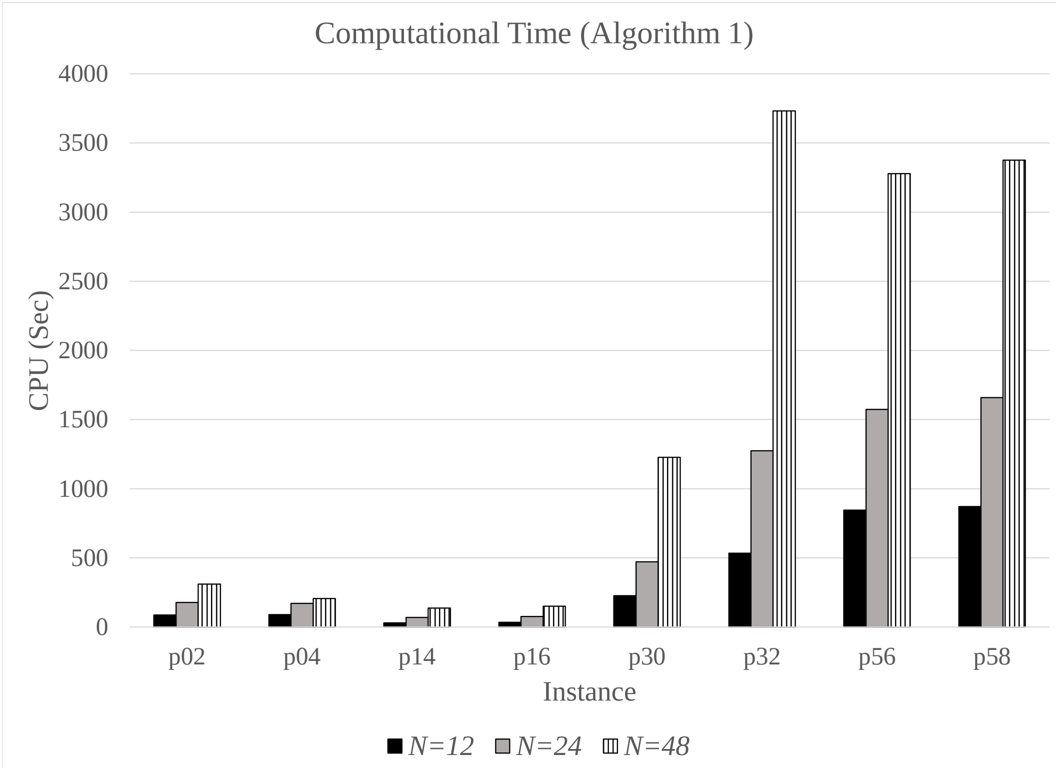


Figure 1: Average Computational Time of the column-and-constraint-generation algorithm.

6.2 Quality of DRO bounds

Next, we investigated the quality of conservative approximations and relaxations proposed in section 5, along with the approximation of the 2-DR-CFLP with its corresponding 1-DR-CFLP, SP and RO formulations. The aim here is to compare these approximations numerically and identify an ordering of lower and upper bounds on the optimal value of 2-DR-CFLP. For each approximation, we report the relative deviation $\left(\frac{V_{approx}-V_{2S}}{V_{2S}}\right)$ from the 2-DR-CFLP optimal value at different values of ε, where V_{approx} is the average optimal value for this approximation and V_{2S} is the average optimal value of the corresponding 2-DR-CFLP, and where averages are taken over the 10 DRO instances tested. Figure 3 shows the results corresponding to four deterministic instances (p02, p14, p30 and p56) of different sizes and with $N = 12$. The effect of sample size on the approximation quality was found to be negligible in all cases, and the effect of facility setup costs was marginal. Thus, results for these deterministic instances were considered representative of the entire test set.

In Figure 3, the curves labeled **SP** and **RO** represent, respectively, the relative deviations of the stochastic programming and robust optimization optimal values from that of 2-DR-CFLP. Although the absolute optimal values in SP and RO do not depend on ε, the relative deviations do since V_{2S} is ε-dependent. When $\varepsilon = 0$, 2-DR-CFLP reduces to a SP problem that considers only the empirical distribution $\widehat{F}_{\xi}^{N}$ and thus the deviation between SP and 2-DR-CFLP equals zero. However, as ε is increased, the two values depart from each other and SP serves only as a lower bounding scheme for DRO. The opposite can be said about RO,

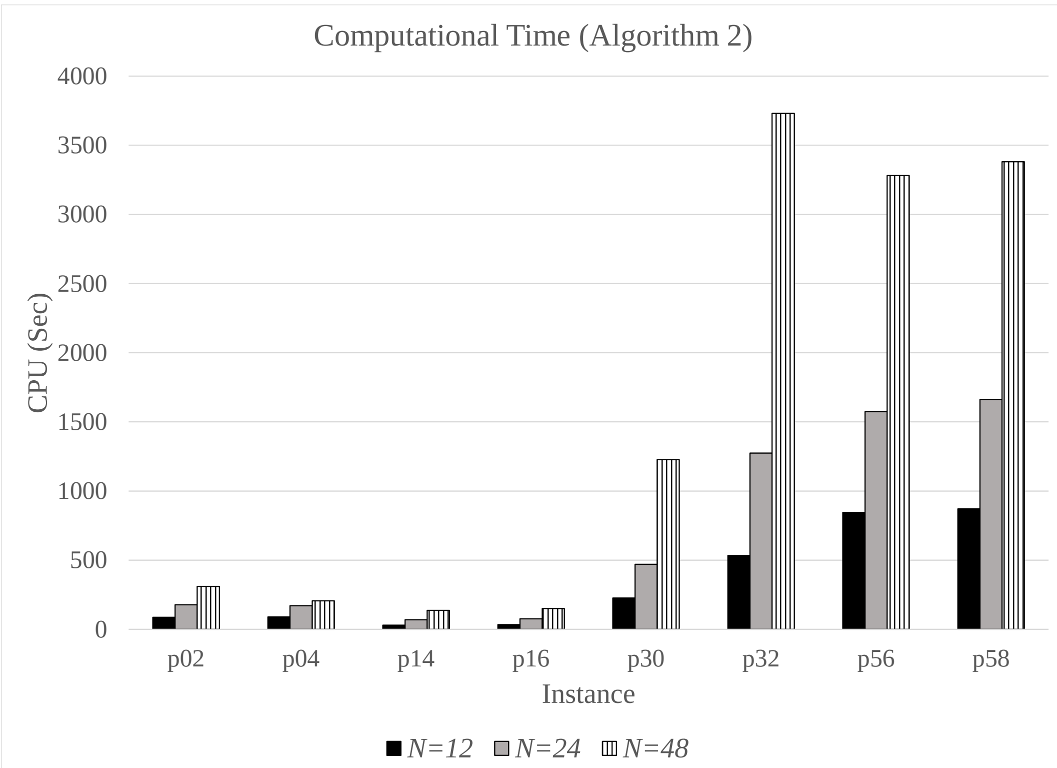
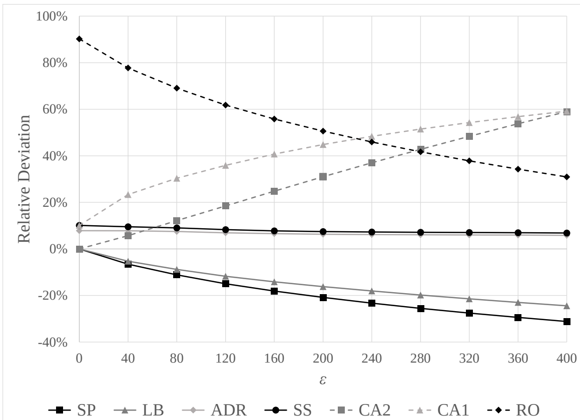
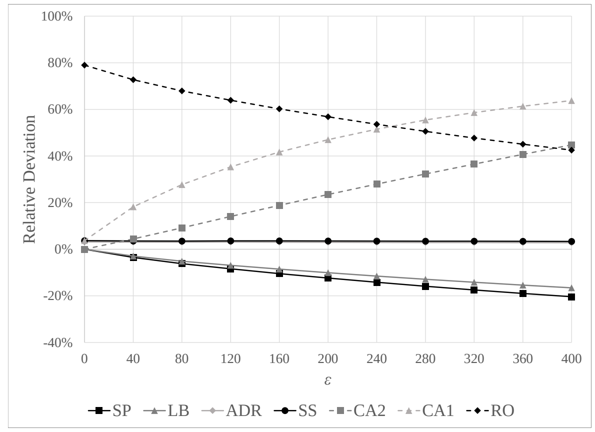


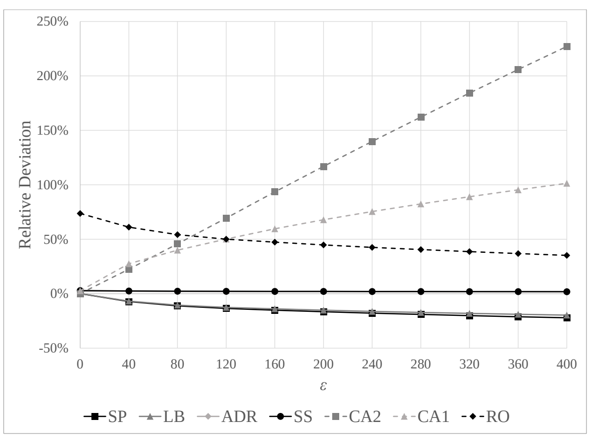
Figure 2: Average Computational Time of the column generation with a support lifting algorithm.



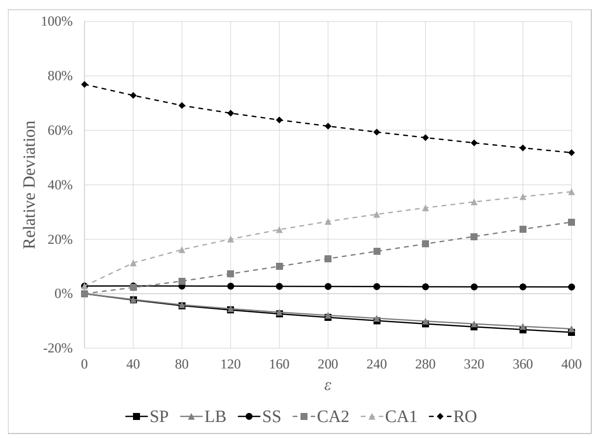
(a)



(b)



(c)



(d)

Figure 3: Relative deviations between the average optimal values of different approximation schemes and the average optimum value of 2-DR-CFLP in DRO instances generated from (a) $p02$, (b) $p14$, (c) $p30$ and (d) $p56$.

which has its largest deviation from 2-DR-CFLP when $\varepsilon = 0$, while this deviation decreases as ε is increased. Although not shown in the figure, when ε becomes large enough, the ambiguity set $\mathcal{D}_\varepsilon(\widehat{F}_\xi^N, \Xi)$ will include probability distributions that place all the probability mass on any one of the scenarios within Ξ, thus the DRO problems reduces to a RO problem and the optimal values of RO and 2-DR-CFLP coincide.

Another lower bound, represented by the curve labeled **LB**, can be obtained through the relaxation scheme described in section 5.3, which restricts the support set of the worst-case distribution to be the support of empirical distribution. Since this is a relaxation, the relative deviation is, similar to that of SP, non-positive. By looking at Figure 3, one can see that this lower bound is only slightly better than that obtained from SP. For example, with $\varepsilon = 400$ in $p02$, the relative deviation for **LB** is -24.4% compared to -31.2% for **SP**. The differences between the two schemes are even smaller in other cases. This result clearly shows that most of the difference between the objective values of DRO and SP is due to the adversary placing probability masses on other realizations than those composing the empirical distribution.

When it comes to upper bounds, it is clear that, except for very small values of ε, the affine decision rule approximation, represented by the curve labeled **ADR**, provides the tightest bound. However, V_{SS}, the optimal value of 1-DR-CFLP, provides a marginally worse (higher) upper bound than V_{ADR}. Hence, it is questionable whether there is much value in solving the ADR approximation instead of the 1-DR-CFLP to obtain an upper bounds and near-optimal solution of the 2-DR-CFLP, given that the former is much more computationally expensive than the latter. For instance, Matlab could not handle the extremely large constraint matrix of the ADR formulation for test instances based on $p56$. The other two conservative approximations, represented by the curves labeled **CA1** and **CA2**, of 1-DR-CFLP and 2-DR-CFLP, respectively are based on support set relaxation (from Ξ to $\mathbb{R}^I$) as explained in section 5.1. One can see that these approximations provide arbitrarily high bounds that can go higher than the RO optimal value when ε becomes large. It is interesting to see that the upper bound obtained from the conservative approximation of 2-DR-CFLP can become worse than what is obtained from the conservative approximation of 1-DR-CFLP despite the fact that 1-DR-CFLP is itself a conservative approximation of 2-DR-CFLP. One explanation is that the quality of the **CA2** bound depends on how the relatively complete recourse problem is converted to a complete recourse one, *i.e.,* going from (4) to (6). Namely, there might be ways of improving it by identifying a set of linear inequalities that describe the convex hull of the vertices of the dual feasible set of problem (6) as discussed in Remark 1 in [1].

Among the different approximation schemes of the 2-DR-CFLP, it seems that solving the corresponding single-stage problem strikes the best balance between computational efficiency and tightness of the bound. The relative deviations of V_{SS} in all DRO instances did not exceed 12%, and were much smaller than that in most cases. In comparison, V_{CA2} provided good bounds only when ε was very small. Nevertheless, an added advantage of this approximation is that it enables the lower bound V_{SP} to be computed at the same time.

An interesting observation to report before closing this part is that, in most cases, applying all of the exact and approximation schemes described earlier (*i.e.,* SP, 2-DR-CFLP, 1-DR-CFLP, RO, ...) on the same DRO instance led to the same optimum first-stage solution x^*. Facility location problems are known, in general, to be relatively insensitive to estimation errors in parameters, especially demand errors (see, for example, [14]), and our results suggest that they are also not very sensitive to the selected risk aversion scheme.

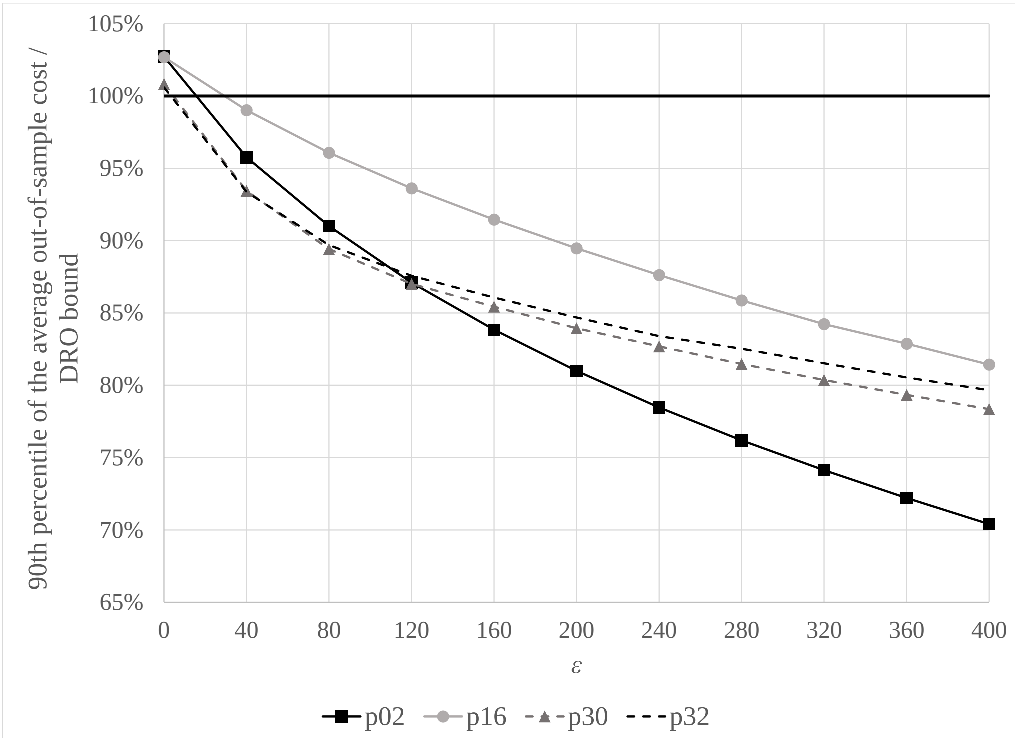


Figure 4: 90th percentile of the ratio between average out-of-sample cost and the DRO bound.

Therefore, our focus in this numerical section was on developing and comparing useful performance bounds instead of merely finding the optimal solutions. Nevertheless, we think that more empirical/theoretical exploration might be needed to confirm or contradict our preliminary finding about scheme insensitivity.

6.3 Out-of-sample Performance Guarantees

Finally, we calculated the out-of-sample average costs corresponding to the optimal solutions of the 2-DR-CFLP and compared them to the DRO bounds calculated using the in-sample data. The purpose of this comparison was to highlight the effect of ε on the performance guarantees of the solutions obtained upon implementing them with new realizations (*i.e.,* out-of-sample data), generated as described earlier. To evaluate the out-of-sample costs, we fix the first-stage variables to the optimal values obtained by solving the 2-DR-CFLP, then re-optimize the second-stage assignments optimally with respect to each of the out-of-sample demand realizations.

Figure 4 shows the 90th-percentiles of the ratios between average out-of-sample costs and their corresponding DRO bounds (*i.e.,* optimal values of the 2-DR-CFLP) in four instance types ($p02$, $p16$, $p30$ and $p32$). One can see that when $\varepsilon = 0$, the 90-percentile of the ratios is greater than one for all instances, meaning that the decision maker will experience post-decision disappointment (*i.e.,* the out-of-sample average cost ends up being higher than the in-sample cost) in more than 10% of the cases. For the in-sample cost to serve as a performance guarantee of the out-of-sample cost in 90% of the cases, ε must be increased such that this ratio equals one at most. As an example, for deterministic instance $p30$, using $\varepsilon = 40$ brings the ratio below one, thus ensures that, in average, out-of-sample costs are less than the V_{2S} with $\varepsilon = 40$ in at least 90% of the cases. The value of ε required to guarantee the performance of a certain percentile is different for each instance, but all the curves are monotonically decreasing. This reflects the fact that one always achieves a stronger probabilistic guarantee by increasing ε in the 2-DR-CFLP.

7 Conclusions

Distributionally Robust Optimization is an attractive alternative to deal with the inherent uncertainty in facility location problems, especially for hedging against distributional ambiguity when decisions have to be based on limited sample data. In this paper, we showed how a data-driven DRO framework with a Wasserstein ambiguity set can be implemented to robustify the classical CFLP. Both single- and two-stage variants of the problem were addressed. For the single-stage case, we provided a reformulation into a mixed integer linear program with a proper selection of the support set and the Wasserstein metric norm. For the two-stage case, we developed two iterative algorithms based on column generation to solve the problem exactly. We also proposed conservative approximations for the single- and two-stage problem based on support set relaxation, a conservative approximation of the two-stage problem using affine decision rules, and a relaxation of the two-stage problem based on support set restriction. All these approximations were reformulated into mixed-integer programs that can be directly handled using commercial solvers.

Our numerical experiments showed that the two exact algorithms proposed for the two-stage problem were able to solve most test instances to optimality in less than one hour per instance. We also found that while affine decision rules provided the tightest conservative approximation for the two-stage problem, the upper bound provided by the single-stage optimal value was almost as good but much easier to obtain. The conservative approximations and relaxations based on support set relaxation and restriction, respectively, were good only for small Wasserstein ball radii. Finally, we empirically investigated the effect of changing the size of the ambiguity set on the out-of-sample performance guarantees for the optimal solutions of the 2-DR-CFLP. Overall, our experiments revealed two interesting empirical observations that could motivate further research. First, the two exact algorithms proposed to solve the two-stage problem have similar computational performances and seem to be equivalent in some sense; and second, the solution of the two-stage stochastic CFLP seems to be relatively insensitive to the risk averse scheme that is implemented.

References

[1] Amir Ardestani-Jaafari and Erick Delage. The value of flexibility in robust location–transportation problems. *Transportation Science*, 52(1):189–209, 2018.

[2] Alper Atamtürk and Muhong Zhang. Two-stage robust network flow and design under demand uncertainty. *Operations Research*, 55(4):662–673, July 2007.

[3] Opher Baron, Joseph Milner, and Hussein Naseraldin. Facility location: A robust optimization approach. *Production and Operations Management*, 20(5):772–785, 2011.

[4] John Gunnar Carlsson, Mehdi Behroozi, and Kresimir Mihic. Wasserstein distance and the distributionally robust TSP. *Operations Research*, 66(6):1603–1624, 2018.

[5] Isabel Correia and Francisco Saldanha da Gama. Facility location under uncertainty. In Gilbert Laporte, Stefan Nickel, and Francisco Saldanha da Gama, editors, *Location Science*, chapter 8, pages 177–203. Springer International Publishing, Oxford, 2015.

[6] M.S. Daskin, S.M. Hesse, and C.S. Revelle. α-reliable p-minimax regret: A new model for strategic facility location modeling. *Location Science*, 5(4):227 – 246, 1997.

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

[8] Desrosiers Jacques Solomon Desaulniers, Guy and Marius M. (Eds.). *Column Generation*. Springer, Boston, MA, 2005.

[9] R. Gao and A. J. Kleywegt. Distributionally Robust Stochastic Optimization with Wasserstein Distance. *ArXiv e-prints*, April 2016.

[10] Joel Goh and Melvyn Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58(4-part-1):902–917, 2010.

[11] Nalan Gülpınar, Dessislava Pachamanova, and Ethem Çanakoğlu. Robust strategies for facility location under uncertainty. *European Journal of Operational Research*, 225(1):21 – 35, 2013.

[12] Trevor S. Hale and Christopher R. Moberg. Location science research: A review. *Annals of Operations Research*, 123(1):21–35, Oct 2003.

[13] Grani A. Hanasusanto and Daniel Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Operations Research*, 66(3):849–869, 2018.

[14] M. John Hodgson. Stability of solutions to the p-median problem under induced data error. *INFOR: Information Systems and Operational Research*, 29(2):167–183, 1991.

[15] Kaj Holmberg, Mikael Rönnqvist, and Di Yuan. An exact algorithm for the capacitated facility location problems with single sourcing. *European Journal of Operational Research*, 113(3):544 – 559, 1999.

[16] R. Horst and N.V. Thoai. Dc programming: Overview. *Journal of Optimization Theory and Applications*, 103:1–43, 1999.

[17] L. V. Kantorovich and G. S. Robinstein. On a space of totally additive functions. *Vestnik Leningradskogo Universiteta*, 13:52–59, 1958.

[18] Esmaeil Keyvanshokooh, Sarah M. Ryan, and Elnaz Kabir. Hybrid robust and stochastic optimization for closed-loop supply chain network design using accelerated benders decomposition. *European Journal of Operational Research*, 249(1):76 – 92, 2016.

[19] Gilbert Laporte, François V. Louveaux, and Luc van Hamme. Exact solution to a location problem with stochastic demands. *Transportation Science*, 28(2):95–103, 1994.

[20] F.V. Louveaux. Stochastic location analysis. *Location Science*, 1(2):127 – 154, 1993.

[21] Fengqiao Luo and Sanjay Mehrotra. Decomposition algorithm for distributionally robust optimization using wasserstein metric with an application to a class of regression models. *European Journal of Operational Research*, 278(1):20 – 35, 2019.

[22] Martin Mevissen, Emanuele Ragnoli, and Jia Yuan Yu. Data-driven distributionally robust polynomial optimization. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 37–45. Curran Associates, Inc., 2013.

[23] Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1):115–166, Sep 2018.

[24] Robert M. Nauss. An improved algorithm for the capacitated facility location problem. *The Journal of the Operational Research Society*, 29(12):1195–1201, 1978.

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

[26] H. Scarf. A min-max solution of an inventory problem. *Studies in The Mathematical Theory of Inventory and Production*, pages 201–209, 1958.

[27] Daniel Serra and Vladimir Marianov. The p-median problem in a changing network: the case of barcelona. *Location Science*, 6(1–4):383 – 394, 1998.

[28] E S Sheppard. A conceptual framework for dynamic location-allocation analysis. *Environment and Planning A*, 6(5):547–564, 1974.

[29] James E. Smith and Robert L. Winkler. The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3):311–322, 2006.

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

[31] Lawrence V. Snyder and Mark S. Daskin. Stochastic p-robust location problems. *IIE Transactions*, 38(11):971–985, 2006.

[32] Wolfram Wiesemann, Daniel Kuhn, and Melvyn Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

[33] David Wozabal. A framework for optimization under ambiguity. *Annals of Operations Research*, 193(1):21–47, 2012.

[34] David Wozabal. Robustifying convex risk measures for linear portfolios: A nonparametric approach. *Operations Research*, 62(6):1302–1315, 2014.

[35] Chenchen Wu, Donglei Du, and Dachuan Xu. An approximation algorithm for the two-stage distributionally robust facility location problem. In David Gao, Ning Ruan, and Wenxun Xing, editors, *Advances in Global Optimization*, pages 99–107. Springer International Publishing, 2015.

[36] B. Zeng and L. Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457 – 461, 2013.

[37] C. Zhao and Y. Guan. Data-driven risk-averse two-stage stochastic program. Technical report, University of Florida, 2014.