Algorithmic Approaches for Identifying the Trade-off between Pessimism and Optimism in a Stochastic Fixed Charge Facility Location Problem

Man Yiu Tsang, Karmel S. Shehadeh

Abstract

We introduce new algorithms to identify the trade-off (TRO) between adopting a distributional belief and hedging against ambiguity when modeling uncertainty in a capacitated fixed charge facility location problem (CFLP). We first formulate a TRO model for the CFLP (TRO-CFLP), which determines the number of facilities to open by minimizing the fixed establishment cost and the maximum expected operational cost evaluated over distributions within a TRO set. This set is defined by an empirical distribution, an ambiguity set, and a TRO parameter that controls the trade-off between solving the TRO-CFLP under the empirical distribution and the worst-case distribution. The TRO-CFLP model enables decision-makers to explore a spectrum of location decisions, from optimistic to conservative. We propose a Spectrum Search Algorithm that identifies the full set of distinct optimal solutions across the TRO parameter space. We also develop a hybrid column-and-constraint generation (hC&CG) algorithm for solving the TRO-CFLP model with a fixed TRO parameter. We employ hC&CG as a subroutine within the Spectrum Search Algorithm. Numerical results demonstrate the computational efficiency of the Spectrum Search Algorithm, the superior performance of hC&CG over state-of-the-art methods, and the practical value of adopting solutions on the TRO-CFLP's spectrum over those obtained using traditional models.

Keywords: Facility location, uncertainty, stochastic programming, distributionally robust optimization

1. Introduction

In this paper, we introduce new algorithmic approaches for investigating the trade-off between adopting a distributional belief for the uncertain demand and hedging against distributional ambiguity in a well-known stochastic capacitated fixed charge facility location problem (CFLP). Given a set of potential locations for the facilities and customer nodes, the CFLP consists of selecting a subset of locations to establish facilities together with an assignment plan that assigns customers' demand to facilities. Facilities are capacitated, and the demand of each customer node is random. The goal is to find location decisions that minimize the fixed costs of establishing facilities

Man Yiu Tsang, Department of Industrial, Manufacturing, and Systems Engineering, Texas Tech University, Lubbock, TX, USA

Karmel S. Shehadeh (corresponding author), Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA, USA

Preprint submitted to Optimization Online

and random operational costs, including transportation costs from customers to open facilities and penalties for unmet demand.

The CFLP lies at the heart of many problems in various application domains, including healthcare, humanitarian logistics, supply chain, energy, and transportation. Classical stochastic programming (SP) models assume that the probability distribution of the demand is known, e.g., following a normal distribution (Ağralı et al., 2012; Shen et al., 2003), Poisson distribution (Lin, 2009; Zhang et al., 2015), or lognormal distribution (Murali et al., 2009). Such models aim to minimize the fixed cost plus the expectation of the random operational costs, where the expectation is evaluated over the assumed distribution. Evaluating the expected operational cost for general distribution can be challenging, as it involves taking multidimensional integrals. Hence, studies often resort to the sample average approximation (SAA) approach, where the true distribution is replaced with the empirical distribution based on a finite set of demand scenarios.

If the probability distribution of the demand is unknown, as is common in many practical applications, we lack an important input parameter for evaluating the cost associated with location decisions. In practice, we often have a small set of historical realizations of the demand or partial information about its distribution (e.g., the mean and range). In this case, we can leverage such partial information to construct a so-called *ambiguity set* of possible distributions against which we want to safeguard. Distributionally robust optimization (DRO) is a class of techniques for modeling uncertainty that employ such a set to hedge against distributional ambiguity. Specifically, a DRO model for the CFLP finds optimal location decisions that minimize the fixed cost plus the maximum expectation of the operational costs, where the expectation is evaluated over the worst-case distribution residing within a pre-defined ambiguity test. Recently, various DRO models have been proposed for the CFLP and its variants (see, e.g., Basciftci et al., 2021; Shehadeh and Sanci, 2021; Liu et al., 2022; Wang et al., 2020).

Solutions to SAA formulations of the CFLP are inherently random and tend to be optimistically biased (Blanchet et al., 2021; Kuhn et al., 2019). In particular, they often display disappointing out-of-sample performance in practice, characterized by significant shortages and high operational costs. Hence, they may display disappointing out-of-sample performance in practice, such as significant shortages and high operational costs. On the other hand, by focusing the optimization on the worst-case distribution, solutions to the DRO model may be pessimistic (conservative). In the context of CFLP, it has been shown that DRO models intend to open more facilities to hedge against ambiguity, leading to a higher fixed cost of opening facilities but lower operational costs than the SP models. Recently, Wang et al. (2023) proposed a Bayesian distributionally robust (BDR) model that balances the trade-off between the specificity of the training data and the robustness of unseen data in machine learning. Tsang and Shehadeh (2025) proposed an application-agnostic trade-off (TRO) approach that offers a middle ground between the optimistic approach (e.g., SAA), which adopts a distributional belief, and the pessimistic (conservative) approach (e.g., DRO), which hedges against

ambiguity. They identify the theoretical properties of the TRO approach and demonstrate how it enables the exploration of a spectrum of decisions, ranging from optimistic to conservative ones. To our knowledge, no one has theoretically investigated such a trade-off in the CFLP. Moreover, the literature lacks computationally efficient algorithms for solving the TRO model and its tailored formulations. Indeed, Tsang and Shehadeh (2025) identified the development of such algorithms as a promising avenue for future research.

1.1. Contributions

In this paper, we introduce algorithmic approaches for identifying the trade-off between following a distributional belief and hedging against distributional ambiguity for modeling demand uncertainty in the CFLP. Our main methodological and computational contributions are as follows.

- We investigate a new trade-off (TRO) model for the CFLP, denoted as the TRO-CFLP model, which determines the optimal number and location of facilities to minimize the fixed costs of opening them and the maximum expected operational cost over distributions residing within a TRO set. The TRO set is characterized by a distributional belief (e.g., the empirical distribution), an ambiguity set representing distributional ambiguity, and a TRO parameter $\theta \in [0, 1]$ that controls the trade-off between solving the problem under a distributional belief and the worst-case distribution within the ambiguity set. We show that by changing θ, decision-makers can obtain a spectrum of solutions to the CFLP, ranging from optimistic to conservative location decisions. Using a general moment-based ambiguity set in the TRO set, we derive an equivalent two-stage robust optimization (RO) formulation of the TRO-CFLP model amenable to decomposition algorithms.
- Solving the TRO-CFLP model for a large set of values of θ is computationally expensive and could potentially yield multiple repeated solutions. To avoid unnecessary waste of computational effort exploring similar solutions, we propose a *Spectrum Search Algorithm* capable of efficiently identifying the entire spectrum of optimal solutions for the TRO-CFLP model. Specifically, this algorithm adaptively identifies the values of the TRO parameter θ associated with each solution on the spectrum based on iteratively improving the gap between lower and upper bounding functions (introduced in Section 5) of the optimal value function of the TRO-CFLP model. We prove the finiteness and correctness of this algorithm. To our knowledge and according to Tsang and Shehadeh (2025), no algorithmic approach has been developed to obtain the full spectrum of distinct solutions for TRO models. Our Spectrum Search Algorithm can be applied to other applications.
- We propose a new hybrid column-and-constraint generation (hC&CG) algorithm for solving the TRO-CFLP reformulation with a fixed value of θ. We utilize this algorithm as a subroutine in the Spectrum Search Algorithm. We prove the finite convergence of our hC&CG method, as well as the classical primal C&CG (pC&CG), dual C&CG (dC&CG), and Benders' decomposition

(BD) algorithms. Our convergence analyses generalize those in Tsang et al. (2023) and Zeng and Zhao (2013) by relaxing the restrictive finiteness or polyhedral assumption on the uncertainty set. While our proposed hC&CG algorithm is motivated by the TRO-CFLP problem, it can be tailored to other applications where the recourse problem has right-hand-side uncertainty and continuous recourse.

- We conduct an extensive computational experiment to analyze the proposed algorithm's computational performance and demonstrate the TRO-CFLP approach's practical benefits. Our results demonstrate the efficiency of the Spectrum Search Algorithm in identifying the full spectrum of optimal solutions for the TRO-CFLP model. Additionally, the results demonstrate the superior computational efficiency of the proposed hC&CG algorithm compared to state-of-the-art decomposition algorithms—BD, pC&CG, and dC&CG—which have been extensively employed in solving two-stage DRO models for facility location and other problems. Our results also highlight the practical benefits of adopting solutions on the TRO-CFLP's spectrum over those obtained using the SAA and DRO approaches. Specifically, we identify practical scenarios where TRO-CFLP solutions lead to lower total costs than those of SAA and DRO models.

1.2. Structure of the Paper

The remainder of the paper is organized as follows. In Section 2, we review the relevant literature. In Section 3, we detail our problem setting. In Section 4, we present and analyze our TRO-CFLP model. In Section 5, we present the Spectrum Search Algorithm and analyze its convergence. In Section 6, we present our decomposition framework for solving the TRO-CFLP model for a given TRO parameter. In Section 7, we present our numerical experiments and insights. We draw conclusions in Section 8.

2. Relevant Literature

For decades, much work has been done on formulating and solving facility location (FL) problems. We refer to Ahmadi-Javid et al. (2017), Daskin (2011), Eiselt and Marianov (2011), Shehadeh and Snyder (2023), and Revelle et al. (2008) for comprehensive surveys on FL models, solution methods, and applications.

Our CFLP belongs to the family of capacitated fixed charge location problems with random demand. This problem and its variants have attracted significant attention from the operations research community. Specifically, two main classes of techniques have been proposed to address uncertainty in the CFLP: stochastic programming (SP) and (distributionally) robust optimization (D/RO). As mentioned earlier, SP is a class of techniques in which one assumes that the distribution of uncertainty is known or that there is high-quality data to estimate it. Sheppard (1974) proposed a scenario-based approach for a plant construction problem. Laporte et al. (1994) proposed an SP model for the CFLP with uncertain demand and an exact solution approach based on Benders' decomposition to solve their model. Santoso et al. (2005) proposed an SP model for a supply chain

network FL problem under cost, demand, and supply uncertainty. They developed an accelerated Benders' decomposition algorithm to solve their model in large-scale settings. Alizadeh et al. (2019) proposed an SP model for the CFLP with heterogeneous demand, assuming the demand follows a Bernoulli distribution. They developed a normal approximation approach to handle heterogeneous demand and proposed a metaheuristic algorithm to solve their model. The SP approach has also been adopted to address uncertainty in various other FL problems, see, e.g., Albareda-Sambola et al. (2017); Bieniek (2015); Conde (2007); Daskin et al. (2002); Louveaux (1986); Shen et al. (2003, 2011); Wagner et al. (2009); Wang and Lee (2015).

As a function of the data, the empirical distribution $\widehat{\mathbb{P}}_N$ (based on N realizations of the demand) is a random object. It follows that optimal decisions obtained from the SP model are also functions of the training data sample. As pointed out by Kuhn et al. (2019), even if one leverages the most sophisticated statistical techniques, the empirical distribution may be significantly different from the true demand distribution. Thus, if $\widehat{\mathbb{P}}_N$ is used instead of the true unknown distribution $\mathbb{P}^*$, the SP model and, thus, solutions to the SP model may inherit any estimation errors in $\widehat{\mathbb{P}}_N$. Thus, optimal decisions to the SP model may display disappointing performance under unseen data when implemented in practice. In the context of the CFLP, such biased location decisions may lead to significant shortages.

Robust optimization (RO) techniques have been proposed to model uncertainty in FL problems, e.g., Baron et al. (2011); Cheng et al. (2021, 2018); Zetina et al. (2017). Distributionally robust optimization (DRO), an approach that dates back to Scarf (1958) and has been of growing interest in recent years, is another approach for modeling uncertainty (Delage and Ye, 2010; Wiesemann et al., 2014). DRO models seek solutions (location decisions) that protect against the worst-case probability distributions over an *ambiguity set*, i.e., a family of possible distributions that share common characteristics. Basciftci et al. (2021) proposed a DRO model for a CFLP problem with decision-dependent uncertain demand. Wang et al. (2020) proposed a DRO model for capacitated and uncapacitated hub-location problems under demand and cost uncertainty. Their ambiguity set includes the demand's mean, absolute dispersion, and support. They reformulate the DRO model as an MILP, which can be solved using off-the-shelf solvers. Their results show superior out-of-sample performance against the classical RO and SP models. Gourtani et al. (2020) proposed two DRO models for the CFLP with uncertain demand using two moment-based ambiguity sets: one based on the mean and covariance matrix of the demand and the other based on the mean. They reformulated the two DRO models as semi-definite and semi-infinite programs and developed numerical schemes to solve the reformulations. Saif and Delage (2021) also studied a DRO version of the CFLP using a Wasserstein ambiguity set. They proposed a primal-based and a dual-based C&CG algorithm for solving the reformulation of their model.

Liu et al. (2019) proposed a DRO model for an emergency medical service station location problem with chance constraints using the ellipsoidal ambiguity set of Delage and Ye (2010). Replacing

the joint chance constraints with individual chance constraints using the Bonferroni approximation, Liu et al. (2019) reformulated their model into a mixed-integer second-order conic program and proposed an outer approximation algorithm to solve the reformulation. Shehadeh and Sanci (2021) proposed a new DRO model for the CFLP with an event-wise ambiguity set that combines two component ambiguity subsets with mean and box support for the demand. As a result, their model reduces to a two-stage adaptive RO problem with binary-box uncertainty. Liu et al. (2022) extended Shehadeh and Sanci (2021) work and proposed an adaptive DRO model for the CFLP using a state-wise ambiguity set. They performed sensitivity analyses of the model by analyzing the impact of change in the ambiguity-set parameters. In addition, they proposed an exact nested Benders' decomposition method to solve their model. Cheng et al. (2024) proposed a DRO model for the CFLP under demand and facility capacity uncertainty with a scenario-wise ambiguity set constructed based on the demand's and capacity's mean and mean absolute deviations. Notably, Shehadeh and Tucker (2022) proposed the first trade-off approach for modeling uncertainty for location and inventory prepositioning of disaster relief supplies. Specifically, the objective of their proposed model is a convex combination of the SP and DRO models' objectives. Using the mean-support ambiguity set, they reformulated the proposed model into an MILP and proposed a decomposition method to solve the reformulation.

Next, we briefly review relevant studies proposing decomposition algorithms for solving two-stage D/RO optimization problems. Thiele et al. (2009) proposed a BD algorithm to solve two-stage RO models with continuous recourse. Later, Zeng and Zhao (2013) introduced the primal C&CG (pC&CG) algorithm to solve two-stage RO models and demonstrated its computational advantages of pC&CG over BD. Since then, pC&CG has become a widely adopted method for solving two-stage D/RO problems across various application domains. Various studies focused on addressing the computational challenges of solving the subproblem in pC&CG. For example, Ayoub and Poss (2016) derived a new formulation of the subproblem under polyhedral uncertainty. Subramanyam (2022) proposed a Lagrangian dual method to solve the subproblem under categorical or binary uncertainty. In various D/RO problems, such as the CFLP, the master problem is a large-scale MILP. Notably, Tsang et al. (2023) is the first study that addressed such a challenge. Specifically, Tsang et al. (2023) proposed an inexact pC&CG algorithm (i-C&CG), which allows solutions to the master problems to be inexact. Tsang et al. (2023)'s numerical results show that i-C&CG outperforms state-of-the-art pC&CG techniques. Recently, Tan et al. (2024) proposed a dual C&CG (dC&CG) algorithm for solving two-stage RO models equipped with a decision-dependent uncertainty set in the context of power systems optimization. Other studies proposing decomposition algorithms for solving various classes of D/RO optimization problems include Arslan and Detienne (2022); Bansal et al. (2018); Kayacık et al. (2025); Tanoumand et al. (2023); Wang et al. (2024). According to recent surveys (Lin et al., 2022; Rahimian and Mehrotra, 2022), and as far as we know, no one has yet proposed algorithmic approaches for solving the TRO model of

Tsang and Shehadeh (2025) and identifying the trade-off between optimism and pessimism. This has been mentioned as a promising future direction in Tsang and Shehadeh (2025).

This paper advances the related literature in several aspects. First, from a modeling perspective, our TRO-CFLP model enables decision-makers to explore a spectrum of solutions to the CFLP, ranging from optimistic to conservative location decisions. Each distinct solution within this spectrum suggests opening a different number of facilities at various locations with different total costs and operational performance. This allows practitioners to select a solution based on their particular settings. Second, from an algorithmic standpoint, we introduce a novel Spectrum Search Algorithm capable of efficiently identifying the entire spectrum of optimal solutions for the TRO-CFLP model. This approach differs from that of Shehadeh and Tucker (2022), who adapted the classical BD algorithm to solve their model with pre-specified values of θ. Third, from another algorithmic perspective, we introduce a new decomposition algorithm (hC&CG). We theoretically and computationally analyze hC&CG relative strength compared to classical dC&CG, pC&CG, and BD algorithms, demonstrating their convergence under general conditions. Our numerical results demonstrate the efficiency of the Spectrum Search Algorithm and the superior computational performance of hC&CG over dC&CG, pC&CG, and BD. Our results also emphasize the practical benefits of adopting solutions on the TRO-CFLP's spectrum. While our proposed algorithms are motivated by CFLP, they can be adapted to a broad range of applications.

3. Problem Setting

Let us now introduce the basic features of the CFLP. We consider a set I of candidate locations for establishing facilities and a set J of customer nodes that generate demand. Each facility's setup or establishing cost is c_i^{f}, for all $i \in I$. The capacity of each facility $i \in I$ is C_i, and the traveling cost from $i \in I$ to $j \in J$ is $t_{i,j}$. The demand d_j at each node $j \in J$ is random. We assume that the support $\mathcal{S} := \{\boldsymbol{d} \in \mathbb{R}^{|J|} \mid \underline{d}_j \leq d_j \leq \overline{d}, \underline{d}_j \geq 0, \overline{d}_j \geq 0, \forall j \in J\}$ of $\boldsymbol{d} = (d_1, \ldots, d_{|J|})^\top$ is known, where $\underline{d}_j$ and $\overline{d}_j$ are respectively the lower and upper bounds on d_j.

As mentioned earlier, the true distribution $\mathbb{P}^\star$ of the demand $\boldsymbol{d}$ is often unknown in most real-life applications. Instead, the decision-maker might have access to a (potentially small) set of N observations $\{\widehat{\boldsymbol{d}}_1, \ldots, \widehat{\boldsymbol{d}}_N\}$ of $\boldsymbol{d}$ (e.g., historical data, a sample generated from a reference distribution, etc.). We define $\widehat{\mathbb{P}}_N = N^{-1} \sum_{i=1}^N \delta_{\widehat{\boldsymbol{d}}_i}$ as the empirical distribution of $\{\widehat{\boldsymbol{d}}_1, \ldots, \widehat{\boldsymbol{d}}_N\}$, where δ_a is the Dirac measure on a. Note that the data $\{\widehat{\boldsymbol{d}}_1, \ldots, \widehat{\boldsymbol{d}}_N\}$ may be inadequate to produce a reliable estimate $\widehat{\mathbb{P}}_N$ of the true unknown distribution $\mathbb{P}^\star$. Other parametric and nonparametric distributions may not perfectly represent $\mathbb{P}^\star$. To model such distributional ambiguity, we consider an *ambiguity set*, which consists of all plausible distributions of $\boldsymbol{d}$. We define $\mathcal{P}(\mathcal{S})$ as the set of probability measures defined on the measurable space $(\mathbb{R}^{|J|}, \mathcal{B})$ induced by $\boldsymbol{d}$ with support $\mathcal{S}$, where $\mathcal{B} = \mathcal{B}(\mathbb{R}^{|J|})$ is the Borel σ-field on $\mathbb{R}^{|J|}$. In addition, we define $\mathcal{F} \subseteq \mathcal{P}(\mathcal{S})$ as the ambiguity set (to be specified later), and $\mathbb{E}_{\mathbb{P}}[\cdot]$ as the expectation with respect to probability measure $\mathbb{P}$.

Due to the random demand and the limited capacity of each facility, there is a possibility

that the opened facilities fail to satisfy all the customers' demands. The penalty of each unit of unsatisfied demand at customer node $j \in J$ is c^{u}_j. We assume that $c^{\mathrm{u}}_j > t_{i,j}$ for all $i \in I$ and $j \in J$. It is easy to verify that the optimal solution will have no facility open without this condition.

Given parameters $(\boldsymbol{C}, \boldsymbol{t}, \boldsymbol{c}^{\mathbf{f}}, \boldsymbol{c}^{\mathbf{u}})$ and sets I and J, we aim to determine the location and the number of facilities to open. Classical approaches for the CFLP seek location decisions that minimize the fixed cost associated with opening facilities plus the expected operational cost (i.e., transportation and unmet demand penalty costs). On the one hand, SP models evaluate the expectation of the operational cost over $\widehat{\mathbb{P}}_N$. Various studies have shown that SP models formulated using an estimated distribution are *optimistically biased* (Mohajerin Esfahani and Kuhn, 2018; Smith and Winkler, 2006; Van Parys et al., 2021). In particular, SP location decisions often perform poorly under unseen data. On the other hand, DRO models evaluate the expectation over the worst-case distribution residing within a pre-defined ambiguity set $\mathcal{F} \subseteq \mathcal{P}(\mathcal{S})$. As mentioned earlier, optimal location decisions of a DRO model are *pessimistically biased*; that is, the realized objective value by such solutions would often be better than the objective value attained by solving the DRO problem. In the next section, we propose a new TRO model that balances the trade-off between the two extremes and identifies its theoretical properties.

4. The Trade-Off Model for the CFLP

In this section, we first introduce the TRO-CFLP model (Section 4.1). Then, we derive equivalent and computationally tractable reformulation of this model (Section 4.2).

4.1. The TRO-CFLP Model

Let us first introduce the decision variables defining our model. We define a binary variable o_i that equals 1 if a facility is established at location $i \in I$ and is 0 otherwise. We define $\mathcal{O} \subseteq \{0,1\}^{|I|}$ as the set of practical constraints related to location decisions $\boldsymbol{o}$. For each $i \in I$ and $j \in J$, we use non-negative continuous variable $x_{i,j}$ to represent the amount of node j's demand satisfied by facility $i \in I$, and non-negative continuous variables u_j to represent the amount of unsatisfied demand of customer node $j \in J$. For each feasible $\boldsymbol{o} \in \mathcal{O}$ and realization of $\boldsymbol{d} = [d_1, \ldots, d_{|J|}]^\top$, we formulate the second-stage (recourse) problem as follows.

$$
\begin{aligned}
Q(\boldsymbol{o}, \boldsymbol{d}) = \ \underset{\boldsymbol{x} \geq \mathbf{0}, \boldsymbol{u} \geq \mathbf{0}}{\text{minimize}} \quad & \sum_{i \in I} \sum_{j \in J} t_{i,j} x_{i,j} + \sum_{j \in J} c^{\mathrm{u}}_j u_j && \text{(1a)} \\
\text{subject to} \quad & \sum_{j \in J} x_{i,j} \leq C_i o_i, \quad \forall i \in I, && \text{(1b)} \\
& \sum_{i \in I} x_{i,j} + u_j \geq d_j, \quad \forall j \in J. && \text{(1c)}
\end{aligned}
$$

Formulation (1a)–(1c) finds optimal second-stage decisions $(\boldsymbol{x}, \boldsymbol{u})$ that minimizes the transportation cost (first term in (1a)) and the penalty cost of unsatisfied demand (second term in (1a)) for a given $\boldsymbol{o} \in \mathcal{O}$ and $\boldsymbol{d}$. Constraints (1b) respect the maximum capacity of each facility, and constraints (1c) compute the unmet demand.

Our TRO-CFLP model can now be stated as follows.

$$\upsilon^\star(\theta) = \underset{\boldsymbol{o}\in\mathcal{O}}{\text{minimize}} \left\{ \sum_{i\in I} c_i^{\text{f}} o_i + \sup_{\mathbb{P}\in\mathcal{F}'(\theta)} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})] \right\}, \tag{2}$$

where we define the TRO ambiguity set $\mathcal{F}'$ as

$$\mathcal{F}'(\theta) = \left\{ (1-\theta)\,\widehat{\mathbb{P}}_N + \theta\,\mathbb{Q} \;\middle|\; \mathbb{Q}\in\mathcal{F} \right\} \tag{3}$$

for some $\theta \in [0,1]$. The TRO set in (3) is characterized by a distributional belief (the empirical distribution $\widehat{\mathbb{P}}_N$), an ambiguity set $\mathcal{F}$ representing distributional ambiguity, and a TRO parameter θ that represents the level of conservatism. In particular, θ controls the trade-off between solving the problem under $\widehat{\mathbb{P}}_N$ and the worst-case distribution within $\mathcal{F}$. When $\theta = 1$, problem (2) reduces to a classical pessimistic DRO problem $\upsilon^\star(1) = \min_{\boldsymbol{o}\in\mathcal{O}} \left\{ \sum_{i\in I} c_i^{\text{f}} o_i + \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})] \right\}$. In contrast, when $\theta = 0$, problem (2) reduces to a classical optimistic SP problem $\upsilon^\star(0) = \min_{\boldsymbol{o}\in\mathcal{O}} \left\{ \sum_{i\in I} c_i^{\text{f}} o_i + \mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\boldsymbol{o},\boldsymbol{d})] \right\}$, where $\mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\boldsymbol{o},\boldsymbol{d})] = (1/N)\sum_{n=1}^{N} Q(\boldsymbol{o},\widehat{\boldsymbol{d}}_n)$. Between the two extremes, $\theta \in (0,1)$ indicates a trade-off between optimistic and pessimistic perceptions of the objective. Hence, by solving problem (2) with distinct values of $\theta \in [0,1]$, we could obtain a range of location decisions from the most optimistic to the most conservative. The following structural properties of the optimal value function $\upsilon^\star$ support this claim.

Theorem 1. *Let $\upsilon^\star(\theta)$ be the optimal value of the TRO-CFLP problem (2) equipped with the TRO set $\mathcal{F}'(\theta)$ defined in (3). The following assertions hold.*

(a) The optimal value function $\upsilon^\star$ is piecewise linear and concave.

(b) If $\widehat{\mathbb{P}}_N \in \mathcal{F}$, the TRO optimal value function $\upsilon^\star$ is a non-decreasing function.

Part (a) of Theorem 1 establishes that $\upsilon^\star$ is piecewise linear and concave, where each linear piece corresponds to a set of optimal solutions. Mathematically, suppose that $\upsilon^\star$ consists of $M-1$ (different) linear pieces defined on the intervals $\{[\theta_{m-1},\theta_m]\}_{m=2}^{M}$. Then, each interval in $\{(\theta_{m-1},\theta_m)\}_{m=2}^{M}$ corresponds to a distinct set of optimal solutions. Part (b) of Theorem 1 suggests that the objective function value of problem (2), and hence the conservatism, increases with θ. Together, parts (a) and (b) indicate that by solving the TRO-CFLP problem (2) for different non-overlapping $\theta \in [0,1]$, we can obtain a spectrum of decisions with optimal values that span $[\upsilon(0),\upsilon(1)]$, i.e., decisions with varying levels of conservatism. As we later show, each distinct solution on the spectrum has a different operational performance and total cost. Practitioners can then choose to implement one of these based on their particular settings. In Section 5, we design an algorithm to obtain this spectrum of optimal solutions. In Section 7, we demonstrate the benefits of adopting solutions on the spectrum over those produced by the traditional SAA and DRO models.

4.2. Reformulation of the TRO-CFLP Model

One can employ any ambiguity set $\mathcal{F}$ in the TRO-CFLP model (2) to represent distributional ambiguity and use the TRO parameter θ to control the degree of conservatism. In this paper, we focus on the following general moment-based ambiguity set. However, our algorithmic strategies can be tailored for other ambiguity sets.

$$\mathcal{F} = \Big\{ \mathbb{P} \in \mathcal{P}(\mathcal{S}) \;\Big|\; \mathbb{E}_{\mathbb{P}}[f_k(\boldsymbol{d})] = \mu_k,\ \forall k \in [K],\ \ \mathbb{E}_{\mathbb{P}}[g_{k'}(\boldsymbol{d})] \leq \sigma_{k'},\ \forall k' \in [K'] \Big\}, \tag{4}$$

where $\{\mu_k\}_{k\in[K]}$ and $\{\sigma_{k'}\}_{k'\in[K']}$ are constants, and $\{f_k : \mathbb{R}^{|J|} \to \mathbb{R}\}_{k\in[K]}$ and $\{g_{k'} : \mathbb{R}^{|J|} \to \mathbb{R}\}_{k'\in[K']}$ are proper, convex, continuous functions on $\boldsymbol{d}$. The first set of constraints in $\mathcal{F}$ captures the moment of $f_k(\boldsymbol{d})$ (e.g., centrality information of $\boldsymbol{d}$), and the second set of constraints captures the upper bounds on the moment of $g_k(\boldsymbol{d})$ (e.g., dispersion of $\boldsymbol{d}$). Ambiguity set $\mathcal{F}$ includes a wide range of well-known moment-based ambiguity sets. For example, when $K = |J|$ with $f_k(\boldsymbol{d}) = d_k$ for $k \in J$ and $K' = 0$, $\mathcal{F}$ reduces to the mean-support (MS) ambiguity set and when $K = |J|$ with $f_k(\boldsymbol{d}) = d_k$ for $k \in J$ and $K' = |J|$ with $g_{k'}(\boldsymbol{d}) = |d_{k'} - \mu_{k'}|$ for $k' \in J$, $\mathcal{F}$ reduces to the mean-absolute-deviation (MAD) ambiguity set.

The TRO-CFLP model in (2) is challenging to solve in its presented form since it involves an inner maximization problem $\sup_{\mathbb{P}\in\mathcal{F}'(\theta)} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o}, \boldsymbol{d})]$. In Proposition 1, we derive an equivalent reformulation of (2) as a single-layer minimization problem.

Proposition 1. *The TRO-CFLP problem (2) with a TRO set $\mathcal{F}'(\theta)$ constructed using the moment-based ambiguity set $\mathcal{F}$ defined in (4) is equivalent to*

$$\underset{\substack{\boldsymbol{o}\in\mathcal{O},\boldsymbol{x}\geq 0,\boldsymbol{u}\geq 0\\ \boldsymbol{\rho},\boldsymbol{\gamma}\geq \mathbf{0},\delta}}{\text{minimize}} \quad \sum_{i\in I} c^{\mathrm{f}}_i o_i + \frac{1-\theta}{N}\sum_{n=1}^{N}\left(\sum_{i\in I}\sum_{j\in J} t_{i,j}x^n_{i,j} + \sum_{j\in J} c^{\mathrm{u}}_j u^n_j\right) + \theta\big(\boldsymbol{\mu}^\top\boldsymbol{\rho} + \boldsymbol{\sigma}^\top\boldsymbol{\gamma} + \delta\big) \tag{5a}$$

$$\text{subject to} \quad \sum_{j\in J} x^n_{i,j} \leq C_i o_i, \quad \forall i \in I,\ n \in [N], \tag{5b}$$

$$\sum_{i\in I} x^n_{i,j} + u_j \geq \widehat{d}^n_j, \quad \forall j \in J,\ n \in [N], \tag{5c}$$

$$\delta \geq Q(\boldsymbol{o}, \boldsymbol{d}) - \sum_{k\in[K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'} g_{k'}(\boldsymbol{d}), \quad \forall \boldsymbol{d} \in \mathcal{S}. \tag{5d}$$

Reformulation (5) is a semi-infinite program due to constraints (5d), which is challenging to solve. This motivates us to propose decomposition algorithms to solve (5), which iteratively solve relaxations of (5) with a finite number of constraints (5d) as detailed in Section 6.

5. The Spectrum Search Algorithm

Solving the TRO-CFLP model for arbitrarily large numbers of pre-specified values of θ is computationally expensive and could potentially yield multiple repeated (similar) solutions (see discussions following Theorem 1). To avoid such unnecessary waste of computational effort exploring similar solutions, in this section, we propose an algorithm called the *Spectrum Search Algorithm*

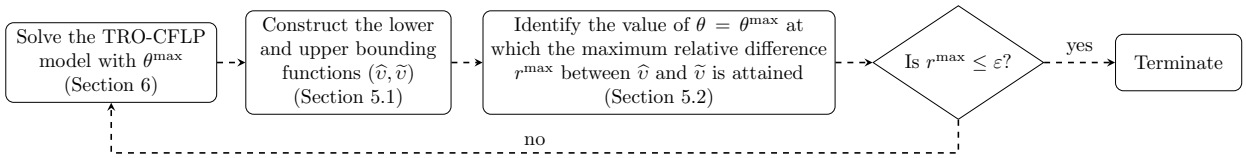


Figure 1: Flow chart of the Spectrum Search Algorithm

that can efficiently identify the full spectrum of distinct optimal solutions for the TRO-CFLP. This algorithm adaptively identifies the values of the TRO parameter θ based on iteratively improving the gap between lower and upper bounding functions $(\widehat{\upsilon}, \widetilde{\upsilon})$ of the optimal value function $\upsilon^\star$ until the maximum relative difference $r^{\max}$ between these two bounding functions (over $\theta \in [0,1]$) is within a pre-specified tolerance $\varepsilon \geq 0$. We use $r^{\max} \leq \varepsilon$ as the termination criterion to ensure that the relative difference between the bounding functions at any $\theta \in [0,1]$ is within ε. This, in turn, implies that optimal solutions in the obtained spectrum are optimal up to a relative gap of ε. Figure 1 provides a high-level summary of the flow of the algorithm. In each iteration, we solve the TRO-CFLP problem (2) with a TRO parameter $\theta^{\max}$ (identified from the previous iteration) to compute lower and upper bounds on $\upsilon^\star(\theta^{\max})$. One can employ the decomposition algorithms proposed in Section 6 or any other algorithm to solve the TRO-CFLP problem (2). Our analyses of the Spectrum Search Algorithm are generic and do not depend on the solution method used to solve the problem. Next, we use the newly obtained lower and upper bounds on $\upsilon^\star(\theta^{\max})$ to construct and update the lower and upper bounding functions $(\widehat{\upsilon}, \widetilde{\upsilon})$ discussed in Section 5.1. Then, we identify a potentially different TRO parameter $\theta^{\max}$ at which the maximum relative difference $r^{\max}$ between the bounding functions is attained. In Section 5.2, we discuss an algorithm to identify such a TRO parameter and obtain the maximum relative difference. If $r^{\max} \leq \varepsilon$, the algorithm terminates; otherwise, it proceeds to the next iteration. In Section 5.3, we present the details of the algorithm and discuss its convergence.

5.1. Lower and Upper Bounding Functions

We first construct a lower bounding function for $\upsilon^\star$. Consider a finite set $\{\theta_m\}_{m=1}^M \subset [0,1]$ with $0 = \theta_1 < \theta_2 < \cdots < \theta_M = 1$. Let $\underline{\upsilon}_m$ and $\overline{\upsilon}_m$ be respectively the lower and upper bounds obtained from solving the TRO-CFLP model with $\theta = \theta_m$ and a relative gap $(\overline{\upsilon}_m - \underline{\upsilon}_m)/\overline{\upsilon}_m \leq \varepsilon$, where $\varepsilon \geq 0$. Using this notation, we define the function $\breve{\upsilon} : [0,1] \to \mathbb{R}$ by linearly interpolating the points $\{(\theta_m, \underline{\upsilon}_m)\}_{m=1}^M$ as follows:

$$\breve{\upsilon}(\theta) = \underline{\upsilon}_1 \mathbf{1}(\theta = 0) + \sum_{m=2}^{M} \left[\frac{\underline{\upsilon}_m - \underline{\upsilon}_{m-1}}{\theta_m - \theta_{m-1}} (\theta - \theta_m) + \underline{\upsilon}_m \right] \mathbf{1}\big(\theta \in (\theta_{m-1}, \theta_m]\big). \tag{6}$$

We define the hypograph of $\breve{\upsilon}$ restricted on the interval $[0,1]$ as $\mathrm{hyp}_{[0,1]}(\breve{\upsilon}) := \{(\theta, z) \mid \theta \in [0,1],\ \breve{\upsilon}(\theta) \geq z\}$, and the convex hull of $\mathrm{hyp}_{[0,1]}(\breve{\upsilon})$ as $\mathrm{conv}\left(\mathrm{hyp}_{[0,1]}(\breve{\upsilon})\right) = \big\{(\theta, \upsilon) = \sum_{k=1}^K \beta_k (\theta_k, \upsilon_k) \mid K \in \mathbb{N},\ (\theta_k, \upsilon_k) \in \mathrm{hyp}_{[0,1]}(\breve{\upsilon}),\ \sum_{k=1}^K \beta_k = 1,\ \{\beta_k\}_{k=1}^K \subset \mathbb{R}_+\big\}$. Using this notation, we define func-

tion $\widehat{\upsilon} : [0,1] \to \mathbb{R}$ as

$$\widehat{\upsilon}(\theta) = \widehat{\upsilon}\big(\theta; \{\theta_m\}_{m=1}^M, \{\underline{\upsilon}_m\}_{m=1}^M\big) := \sup\big\{z \;\big|\; (\theta, z) \in \operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)\big\} \tag{7a}$$

$$= \inf\big\{g(\theta) \;\big|\; g : [0,1] \to \mathbb{R} \text{ is concave, } g \geq \breve{\upsilon}\big\}. \tag{7b}$$

Note that function $\widehat{\upsilon}$ defined in (7a) is the concave envelope of $\breve{\upsilon}$, which is equivalent to the pointwise minimum of all concave function g that dominates $\breve{\upsilon}$ on $[0,1]$ defined in (7b) (see, e.g., Corollary 17.1.5 of Rockafellar, 1997). That is, $\widehat{\upsilon}$ is the tightest concave function we can construct based on the given information $\{\theta_m\}_{m=1}^M$ and $\{\underline{\upsilon}_m\}_{m=1}^M$. In Proposition 2, we identify structural properties of function $\widehat{\upsilon}$ and show that it is a lower bounding function of $\upsilon^\star$.

Proposition 2. *Consider the finite set of points $\{(\theta_m, \underline{\upsilon}_m)\}_{m=1}^M$, where $0 = \theta_1 < \theta_2 < \ldots < \theta_M = 1$ and $\underline{\upsilon}_m$ is the lower bound obtained from solving the TRO-CFLP with $\theta = \theta_m$. The following assertions hold.*

(a) The function $\widehat{\upsilon}$ defined in (7a) is piecewise linear and concave with $\widehat{\upsilon} \leq \upsilon^\star$. In particular, $\widehat{\upsilon}$ is the function linearly interpolating the extreme points of $\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)$.

(b) Suppose $\widehat{\mathbb{P}}_N \in \mathcal{F}$. Consider the modified set $\{(\theta_m, \underline{\upsilon}'_m)\}_{m=1}^M$, where $\underline{\upsilon}'_m := \max_{m' \leq m}\{\underline{\upsilon}_{m'}\}$ for all $m \in [M]$. The function $\widehat{\upsilon}$ constructed using this set in (7a) is non-decreasing, piecewise linear, and concave with $\widehat{\upsilon} \leq \upsilon^\star$.

Part (a) of Proposition 2 suggests that we can construct $\widehat{\upsilon}$ by identifying the extreme points of $\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big) \subseteq [0,1] \times \mathbb{R}$. One could apply any numerical algorithm for finding two-dimensional convex hulls to obtain these extreme points and construct the lower bounding function $\widehat{\upsilon}$. In Appendix B.1, we present a gift-wrapping algorithm tailored to our context to obtain the extreme points of $\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big) \subseteq [0,1] \times \mathbb{R}$ (Jarvis, 1973). In particular, for any $m \in [M]$ such that $\underline{\upsilon}_m < \widehat{\upsilon}(\theta_m)$, the algorithm improves the lower bounding value $\underline{\upsilon}_m$ to $\widehat{\upsilon}(\theta_m)$. Therefore, we can construct $\widehat{\upsilon}$ by linearly interpolating the updated points $\{(\theta_m, \underline{\upsilon}_m)\}_{m=1}^M$. Part (b) of Proposition 2 shows that if $\widehat{\mathbb{P}}_N \in \mathcal{F}$, one can construct the lower bounding function $\widehat{\upsilon}$ using the values $\{\underline{\upsilon}'_m\}_{m=1}^M$ instead of $\{\underline{\upsilon}_m\}_{m=1}^M$, where $\underline{\upsilon}'_m := \max_{m' \leq m}\{\underline{\upsilon}_{m'}\}$ for all $m \in [M]$. By construction, $\underline{\upsilon}'_m \geq \underline{\upsilon}_m$ for all $m \in [M]$. It follows that constructing $\widehat{\upsilon}$ using $\{\underline{\upsilon}'_m\}_{m=1}^M$ gives a tighter lower bounding function.

Next, we construct an upper bounding function for $\upsilon^\star$. Let $\widehat{\boldsymbol{o}}_m$ be the best feasible solution obtained from solving the TRO-CFLP problem (2) with $\theta = \theta_m$ and relative gap $\varepsilon > 0$ for $m \in [M]$. Let $\widehat{\varphi}_m^{\mathrm{SP}} := \mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\widehat{\boldsymbol{o}}_m, \boldsymbol{d})]$ be the expected recourse (under the empirical distribution) and $\widehat{\varphi}_m^{\mathrm{DRO}} := \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\widehat{\boldsymbol{o}}_m, \boldsymbol{d})]$ be the worst-case recourse evaluated at $\widehat{\boldsymbol{o}}_m$ for $m \in [M]$. It is easy to verify that the function

$$\widetilde{\upsilon}^{\mathrm{exact}}(\theta) = \widetilde{\upsilon}^{\mathrm{exact}}\big(\theta; \{\theta_m\}_{m=1}^M, \{\widehat{\boldsymbol{o}}_m\}_{m=1}^M\big) := \min_{m \in [M]} \left\{ \sum_{i \in I} c_i^{\mathrm{f}}(\widehat{\boldsymbol{o}}_m)_i + (1-\theta)\widehat{\varphi}_m^{\mathrm{SP}} + \theta \widehat{\varphi}_m^{\mathrm{DRO}} \right\}$$

is an upper bounding function for $\upsilon^\star$, where $(\widehat{\boldsymbol{o}}_m)_i$ is the ith entry of $\widehat{\boldsymbol{o}}_m$. One can efficiently compute $\widehat{\varphi}_m^{\mathrm{SP}}$ by solving the second-stage linear program (LP) $\mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\widehat{\boldsymbol{o}}_m, \boldsymbol{d})]$ using off-the-shelf solvers. In contrast, to obtain $\widehat{\varphi}_m^{\mathrm{DRO}}$, we need to solve the max-min problem $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\widehat{\boldsymbol{o}}_m, \boldsymbol{d})]$, for example, using the decomposition algorithms proposed in Section 6 with $\boldsymbol{o}$ fixed to $\widehat{\boldsymbol{o}}_m$ and $\theta = 1$. Thus, evaluating $\widehat{\varphi}_m^{\mathrm{DRO}}$ for different m could be computationally expensive depending on the choice of the ambiguity set $\mathcal{F}$. This motivates us to consider an approximation of $\widetilde{\upsilon}^{\mathrm{exact}}$ by replacing $\{\widehat{\varphi}_m^{\mathrm{DRO}}\}_{m=1}^M$ with some upper bounds $\{\acute{\varphi}_m^{\mathrm{DRO}}\}_{m=1}^M$ that are easier to compute. For example, assume that the objective value of the TRO-CFLP model evaluated at $\widehat{\boldsymbol{o}}_m$, i.e., $f_m(\widehat{\boldsymbol{o}}_m) := \sum_{i\in I} c_i^{\mathrm{f}}(\widehat{\boldsymbol{o}}_m)_i + (1-\theta_m)\mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\widehat{\boldsymbol{o}}_m, \boldsymbol{\xi})] + \theta_m \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\widehat{\boldsymbol{o}}_m, \boldsymbol{\xi})]$, is always less than or equal to the upper bound $\overline{\upsilon}_m$. This is true, for example, when solving the TRO-CFLP model using Algorithm 3 introduced in Section 6. Then, we can obtain an upper bound $\acute{\varphi}_m^{\mathrm{DRO}}$ of $\widehat{\varphi}_m^{\mathrm{DRO}}$ as follows:

$$\acute{\varphi}_m^{\mathrm{DRO}} := \frac{1}{\theta_m}\left[\overline{\upsilon}_m - \sum_{i\in I} c_i^{\mathrm{f}}(\widehat{\boldsymbol{o}}_m)_i - (1-\theta_m)\widehat{\varphi}_m^{\mathrm{SP}}\right] \geq \frac{1}{\theta_m}\left[f_m(\widehat{\boldsymbol{o}}_m) - \sum_{i\in I} c_i^{\mathrm{f}}(\widehat{\boldsymbol{o}}_m)_i - (1-\theta_m)\widehat{\varphi}_m^{\mathrm{SP}}\right] = \widehat{\varphi}_m^{\mathrm{DRO}} \tag{8}$$

for all $m \in \{2, \ldots, M\}$. The upper bound $\acute{\varphi}_m^{\mathrm{DRO}}$ in (8) can be easily computed after solving the TRO-CFLP model. Using $\{\acute{\varphi}_m^{\mathrm{DRO}}\}_{m=1}^M$, we define the function $\widetilde{\upsilon} : [0,1] \to \mathbb{R}$ as

$$\widetilde{\upsilon}(\theta) = \widetilde{\upsilon}\big(\theta; \{\theta_m\}_{m=1}^M, \{\widehat{\boldsymbol{o}}_m\}_{m=1}^M, \{\acute{\varphi}_m^{\mathrm{DRO}}\}_{m=1}^M\big) := \min_{m\in[M]} \left\{ \sum_{i\in I} c_i^{\mathrm{f}}(\widehat{\boldsymbol{o}}_m)_i + (1-\theta)\widehat{\varphi}_m^{\mathrm{SP}} + \theta\acute{\varphi}_m^{\mathrm{DRO}} \right\}. \tag{9}$$

In Proposition 3, we identify structural properties of function $\widetilde{\upsilon}$ and show that it is an upper bounding function of $\upsilon^\star$.

Proposition 3. *Consider the function $\widetilde{\upsilon}$ defined in (9) for any given finite set of $\{\theta_m\}_{m=1}^M \subset [0,1]$, set of feasible solutions $\{\widehat{\boldsymbol{o}}_m\}_{m=1}^M$, and set of upper bounds $\{\acute{\varphi}_m^{\mathrm{DRO}}\}_{m=1}^M$. The following assertions hold.*

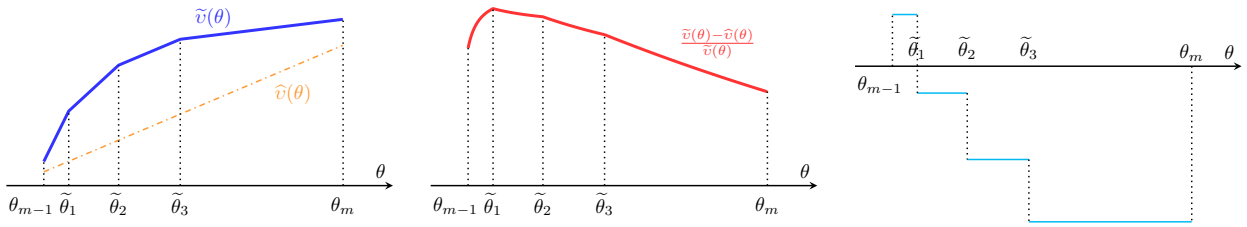
(a) $\widetilde{\upsilon}$ is piecewise linear and concave with $\widetilde{\upsilon} \geq \upsilon^\star$.

(b) If $\widehat{\mathbb{P}}_N \in \mathcal{F}$, then $\widetilde{\upsilon}$ is non-decreasing.

From Propositions 2 and 3, we know that for any given finite set $\{\theta_m\}_{m=1}^M \subset [0,1]$, set of feasible solutions $\{\widehat{\boldsymbol{o}}_m\}_{m=1}^M$, and set of upper bounds $\{\acute{\varphi}_m^{\mathrm{DRO}}\}_{m=1}^M$, we have $\widehat{\upsilon}(\theta) \leq \upsilon^\star(\theta) \leq \widetilde{\upsilon}(\theta)$. In the next section, we present an algorithm to obtain the maximum relative difference of a given pair of lower and upper bounding functions and the corresponding value of θ where this maximum is achieved.

5.2. The Maximum Relative Difference Between the Bounding Functions

Recall that in each iteration of the Spectrum Search Algorithm, we identify the TRO parameter $\theta^{\max}$ at which the *maximum relative difference* between a given pair of the bounding functions $(\widehat{\upsilon}, \widetilde{\upsilon})$, denoted as $r^{\max}$, is attained. By solving the TRO-CFLP model with this new $\theta^{\max}$, we could strictly improve the bounding function(s) in the next iteration. To this end, in this section, we present



(a) $\widetilde{v}$ and $\widehat{v}$ (b) Relative difference (c) Determinant

Figure 2: An illustration of the upper bounding function $\widetilde{v}$ and lower bounding function $\widehat{v}$ on $[\theta_{m-1}, \theta_m]$, the relative difference between $\widetilde{v}$ and $\widehat{v}$, and the determinant value of the relative difference. The upper bounding function $\widetilde{v}$ consists of $K = 4$ pieces with kink points $\{\widetilde{\theta}_1, \widetilde{\theta}_2, \widetilde{\theta}_3\}$.

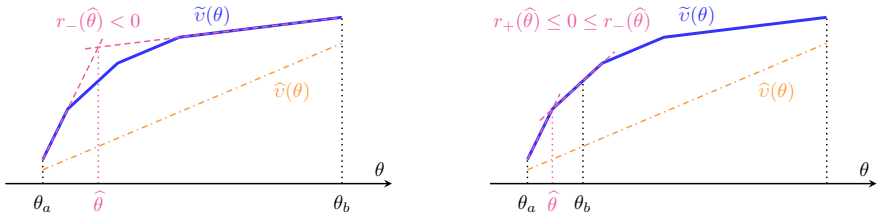
an algorithm to identify $r^{\max}$ and the associated $\theta^{\max}$. We define the maximum relative difference between $\widehat{v}$ and $\widetilde{v}$ as $r^{\max} := \||\widetilde{v} - \widehat{v}|/\widetilde{v}\|_\infty = \sup_{\theta\in[0,1]} \{|\widetilde{v}(\theta) - \widehat{v}(\theta)|/\widetilde{v}(\theta)\}$.

Recall that, on a given interval $[\theta_{m-1}, \theta_m]$, the lower bounding function $\widehat{v}$ is linear and concave, whereas the upper bounding function $\widetilde{v}$ is piecewise linear and concave; see Figure 2a. Each linear piece of $\widetilde{v}$ changes direction at specific points called *kink points* or *breakpoints*. Hence, the relative difference between $\widetilde{v}$ and $\widehat{v}$ can change most significantly at the kink points of $\widetilde{v}$ (i.e., where the slope of $\widetilde{v}$ changes); see Figure 2b for an example. Let p_m be the slope of $\widehat{v}$ on $[\theta_{m-1}, \theta_m]$. Also, let $s_-(\theta)$ and $s_+(\theta)$ be the left and right derivatives of $\widetilde{v}$, respectively; see Appendix B.2. Accordingly, we define $r_-(\theta) = -p_m\widetilde{v}(\theta) + s_-(\theta)\widehat{v}(\theta)$ and $r_+(\theta) = -p_m\widetilde{v}(\theta) + s_+(\theta)\widehat{v}(\theta)$ as the *left* and *right determinants* of the relative difference at θ, respectively. Assume that $\widehat{v} > 0$ and let $\widetilde{\theta}_0 = \theta_{m-1}$ and $\{\widetilde{\theta}_1, \ldots, \widetilde{\theta}_{K-1}\} \subset (\theta_{m-1}, \theta_m)$ be the kink points of $\widetilde{v}$ on (θ_{m-1}, θ_m) with $\widetilde{\theta}_1 < \widetilde{\theta}_2 < \cdots < \widetilde{\theta}_{K-1}$. Note that $r_+(\widetilde{\theta}_{j-1}) = r'_-(\widetilde{\theta}_j)$ for all $j \in [K-1]$; see Figure 2c for an illustration. In addition, the value of the determinant $r_+(\theta)$ at the kink points of $\widetilde{v}$ is decreasing, i.e., $r_+(\widetilde{\theta}_j) < r_+(\widetilde{\theta}_{j-1})$ for all $j \in [K-1]$; see Lemma 2 in Appendix A.5. Lemma 3 in Appendix A.5 suggests that the value of the relative difference is increasing/remains unchanged/is decreasing on $[\widetilde{\theta}_{j-1}, \widetilde{\theta}_j]$ if and only if $r_+(\widetilde{\theta}_{j-1})$ is greater than 0/equal to 0/less than 0. Together, Lemmas 2 and 3 provide a mechanism to identify the value of θ at which the maximum relative difference $\theta_m^{\max}$ is attained on $[\theta_{m-1}, \theta_m]$. This involves analyzing the sign of the determinants $r_-(\theta)$ and $r_+(\theta)$ on each linear piece of $\widetilde{v}$. We formally prove this in the following proposition.

Proposition 4. *Assume that $\widehat{v} > 0$. Consider the interval $[\theta_{m-1}, \theta_m]$ for $m \in \{2, \ldots, M\}$. Let $\widetilde{\theta}_0 = \theta_{m-1}$ and $\{\widetilde{\theta}_1, \ldots, \widetilde{\theta}_{K-1}\} \subset (\theta_{m-1}, \theta_m)$ be the kink points of the upper bounding function $\widetilde{v}$ on (θ_{m-1}, θ_m) with $\widetilde{\theta}_1 < \widetilde{\theta}_2 < \cdots < \widetilde{\theta}_{K-1}$.*

(a) If $r_+(\theta_{m-1}) \le 0$, the maximum relative difference between $\widetilde{v}$ and $\widehat{v}$ on $[\theta_{m-1}, \theta_m]$ is attained at θ_{m-1}.

(b) If $r_-(\theta_m) \ge 0$, the maximum relative difference between $\widetilde{v}$ and $\widehat{v}$ on $[\theta_{m-1}, \theta_m]$ is attained at θ_m.



(a) Iteration 1 (b) Iteration 2

Figure 3: An illustration of the iterative procedure (step 2) in Algorithm 1. In each iteration, the plot shows the search interval (θ_a, θ_b), the value of $\widehat{\theta}$, and one of the three conditions satisfied in step 2b; see Figure 2c for the values of the determinants.

(c) If there exists $j' \in [K-1]$ such that j' is the first index for which $r_+(\widetilde{\theta}_{j'}) \le 0$, the maximum relative difference between $\widetilde{v}$ and $\widehat{v}$ on $[\theta_{m-1}, \theta_m]$ is attained at the kink point $\widetilde{\theta}_{j'}$.

In Algorithm 1, leveraging the fact that the lower bounding function $\widehat{v}$ is linear on each interval $[\theta_{m-1}, \theta_m]$, we first obtain the set of maximizers $\{\theta_m^{\max}\}_{m=1}^M$ and the set of maximum relative differences $\{r_m^{\max}\}_{m=1}^M$ on $\{[\theta_{m-1}, \theta_m]\}_{m=1}^M$. Since $\cup_{m=1}^M [\theta_{m-1}, \theta_m] = [0, 1]$, we can then obtain the maximum relative difference $r^{\max}$ and thus identify $\theta^{\max}$. Algorithm 1 proceeds as follows. In step 1, we compute the right determinant at the left endpoint $r_+(\theta_{m-1})$ and the left determinant at the right endpoint $r_-(\theta_m)$. If $r_+(\theta_{m-1}) \le 0$, it follows from part (a) of Proposition 4 that $r_m^{\max}$ is attained at $\theta_m^{\max} = \theta_{m-1}$. If $r_-(\theta_m) \ge 0$, it follows from part (b) of Proposition 4 that $r_m^{\max}$ is attained at $\theta_m^{\max} = \theta_m$. If $r_+(\theta_{m-1}) > 0$ and $r_-(\theta_m) < 0$, part (c) of Proposition 4 shows that $\theta_m^{\max}$ is attained at some kick points on (θ_{m-1}, θ_m). To search for $\theta_m^{\max}$, we proceed to step 2. In this step, we first initialize the interval (θ_a, θ_b) as (θ_{m-1}, θ_m) and iteratively refine the search region (θ_a, θ_b) until we identify $\theta_m^{\max}$. In step 2a, we compute a new value $\widehat{\theta} \in (\theta_a, \theta_b)$ using (10) and the left and right determinants at $\theta = \widehat{\theta}$. (Here, $\widehat{\theta}$ is the intersection point of the two linear pieces $s_+(\theta_a)\cdot(\theta-\theta_a)+\widetilde{v}(\theta_a)$ and $s_-(\theta_b)\cdot(\theta-\theta_b)+\widetilde{v}(\theta_b)$ in $\widetilde{v}$). Then, in step 2b, we check if we have identified $\theta_m^{\max}$ or need to refine the search region. Specifically, if $r_+(\widehat{\theta}) \le 0 \le r_-(\widehat{\theta})$, then $\theta_m^{\max} = \widehat{\theta}$ by part (c) of Proposition 4. Otherwise, we refine the search region (θ_a, θ_b) and return to step 2a. We illustrate this iterative procedure in Figure 3. Finally, using the identified set of maximizers $\{\theta_m^{\max}\}_{m=1}^M$, we compute the maximum relative differences $\{r_m^{\max} := |\widetilde{v}(\theta_m^{\max}) - \widehat{v}(\theta_m^{\max})|/\widetilde{v}(\theta_m^{\max})\}_{m=1}^M$ over the intervals $\{[\theta_{m-1}, \theta_m]\}_{m=1}^M$. Then, we search for an index $m' \in \{2, \ldots, M\}$ such that $r_{m'}^{\max} \ge r_m^{\max}$ for all $m \in \{2, \ldots, M\}$. The algorithm return the maximum relative difference $r^{\max} = r_{m'}^{\max}$ and the TRO parameter $\theta^{\max} = \theta_{m'}^{\max}$ at which $r^{\max}$ is attained.

In Theorem 2, we prove the finiteness and correctness of Algorithm 1.

Theorem 2. *Let r^{max} and θ^{max} be the outputs of Algorithm 1 for a given pair of bounding functions $(\widehat{v}, \widetilde{v})$. Then, r^{max} is the maximum relative difference between the bounding functions and θ^{max} is the value of the TRO parameter at which the r^{max} is attained. In addition, Algorithm 1 terminates*

Algorithm 1: An algorithm to obtain $r^{\max} = \|\|\widetilde{v} - \widehat{v}|/\widetilde{v}\|_\infty$ and the value of $\theta^{\max}$ at which $r^{\max}$ is attained

Input: A set $\{\theta_m\}_{m=1}^M$ of θ values, upper bounding function $\widetilde{v}$, slopes p_m of the lower bounding function $\widehat{v}$ on each interval $[\theta_{m-1}, \theta_m]$ for $m \in \{2, \ldots, M\}$

foreach $m \in \{2, \ldots, M\}$ **do**

1. Initial Check.

a. Compute $r_+(\theta_{m-1}) = -p_m\widetilde{v}(\theta_{m-1}) + s_+(\theta_{m-1})\widehat{v}(\theta_{m-1})$ and $r_-(\theta_m) = -p_m\widetilde{v}(\theta_m) + s_-(\theta_m)\widehat{v}(\theta_m)$.

b. If $r_+(\theta_{m-1}) \leq 0$, set $\theta_m^{\max} \leftarrow \theta_{m-1}$.

c. If $r_-(\theta_m) \geq 0$, set $\theta_m^{\max} \leftarrow \theta_m$.

2. Iterative Procedure.

if $r_+(\theta_{m-1}) > 0$ **and** $r_-(\theta_m) < 0$ **then**

Initialize $\theta_a \leftarrow \theta_{m-1}$ and $\theta_b \leftarrow \theta_m$.

a. Obtain a new value of $\theta \in (\theta_a, \theta_b)$**.**

- Compute $\widehat{\theta}$ as

$$\widehat{\theta} = \frac{\left[\widetilde{v}(\theta_b) - \theta_b s_-(\theta_b)\right] - \left[\widetilde{v}(\theta_a) - \theta_a s_+(\theta_a)\right]}{s_+(\theta_a) - s_-(\theta_b)}. \tag{10}$$

- Compute $r_+(\widehat{\theta}) = -p_m\widetilde{v}(\widehat{\theta}) + s_+(\widehat{\theta})\widehat{v}(\widehat{\theta})$ and $r_-(\widehat{\theta}) = -p_m\widetilde{v}(\widehat{\theta}) + s_-(\widehat{\theta})\widehat{v}(\widehat{\theta})$.

b. Update (θ_a, θ_b)**.**

- If $r_+(\widehat{\theta}) \leq 0 \leq r_-(\widehat{\theta})$, set $\theta_m^{\max} \leftarrow \widehat{\theta}$.
- If $r_+(\widehat{\theta}) > 0$, set $\theta_a \leftarrow \widehat{\theta}$ and return to **step 2a**.
- If $r_-(\widehat{\theta}) < 0$, set $\theta_b \leftarrow \widehat{\theta}$ and return to **step 2a**.

end

end

Set $r_m^{\max} = |\widetilde{v}(\theta_m^{\max}) - \widehat{v}(\theta_m^{\max})|/\widetilde{v}(\theta_m^{\max})$ for all $m \in \{2, \ldots, M\}$.

Search for an index $m' \in \{2, \ldots, M\}$ such that $r_{m'}^{\max} \geq r_m^{\max}$ for all $m \in \{2, \ldots, M\}$.

Return the maximum relative difference $r^{\max} = r_{m'}^{\max}$ and the TRO parameter $\theta^{\max} = \theta_{m'}^{\max}$.

in a finite number of iterations.

Theorem 2 shows that we can apply Algorithm 1 to obtain the maximum relative difference $r^{\max}$ between the lower and upper bounding functions and the TRO parameter $\theta^{\max}$ at which $r^{\max}$ is attained. It follows that one can strictly improve the bounding function(s) by solving the TRO-CFLP problem with $\theta^{\max}$. These observations motivate our Spectrum Search Algorithm described in the next section, which finds $\theta^{\max}$ iteratively until $r^{\max} \leq \varepsilon$.

5.3. Steps of the Spectrum Search Algorithm

Algorithm 2 summarizes the steps of our Spectrum Search Algorithm. We initialize the algorithm with a set of TRO parameters $\Theta_1 = \{0, 1\}$. Each iteration ℓ starts with the set Θ_ℓ and proceeds as follows. In step 1, we solve the TRO-CFLP problem (2) with $\theta \in \overline{\Theta}$ to within a relative optimality gap ε and record the following: the lower and upper bounds $\big(\underline{v}(\theta), \overline{v}(\theta)\big)$ of the true

optimal value $\upsilon^\star(\theta)$ and the best feasible solution $\widehat{\boldsymbol{o}}(\theta)$ for $\theta \in \overline{\Theta}$. At iteration $\ell = 1$, we have $\overline{\Theta} = \{0, 1\}$; at iteration $\ell > 1$, we have $\overline{\Theta} = \{\theta^{\max}\}$, where $\theta^{\max}$ is identified in iteration $\ell - 1$. Then, we compute the expected recourse $\widehat{\varphi}^{\mathrm{SP}}(\theta) = \mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\widehat{\boldsymbol{o}}(\theta), \boldsymbol{d})]$ for all $\theta \in \overline{\Theta}$. In step 1d, at iteration $\ell > 1$, we check whether $\widehat{\boldsymbol{o}}(\theta^{\max}) \in \{\widehat{\boldsymbol{o}}(\theta) \mid \theta \in \Theta_{\ell-1}\}$, i.e., whether the solution we found in iteration ℓ has been obtained in previous iterations. If $\widehat{\boldsymbol{o}}(\theta^{\max}) \in \{\widehat{\boldsymbol{o}}(\theta) \mid \theta \in \Theta_{\ell-1}\}$, we evaluate the worst-case recourse $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\widehat{\boldsymbol{o}}(\theta^{\max}), \boldsymbol{d})]$ exactly and set $\acute{\varphi}^{\mathrm{DRO}}(\theta)$ to this value; otherwise, we set $\acute{\varphi}^{\mathrm{DRO}}(\theta)$ as an upper bound of the worst-case recourse, using, e.g., (8).

In step 2, we update the lower bounding values as follows. First, we sort and enumerate the set Θ_ℓ as $\{\theta_m\}_{m=1}^M$ such that its elements satisfy $0 = \theta_1 < \theta_2 < \cdots < \theta_M = 1$, where $M = |\Theta_\ell|$. Then, we set $\underline{\upsilon}_m = \underline{\upsilon}(\theta_m)$ and $\overline{\upsilon}_m = \overline{\upsilon}(\theta_m)$ for all $m \in [M]$. If $\widehat{\mathbb{P}}_N \in \mathcal{F}$, it follows from part (b) of Proposition 2 that we can update the lower bounding values as $\underline{\upsilon}_m = \max_{m' \le m}\{\underline{\upsilon}_{m'}\}$ for all $m \in [M]$. Next, following the discussions after Proposition 2, we apply a tailored gift wrapping algorithm (Algorithm 4 in Appendix B.1) to update $\{\underline{\upsilon}_m\}_{m=1}^M$. In step 3, we update the lower bounding function $\widehat{\upsilon}^\ell$ by linearly interpolating the points $\{(\theta_m, \underline{\upsilon}_m)\}_{m=1}^M$, and the upper bounding function $\widetilde{\upsilon}^\ell$ as defined in (9). Finally, in step 4, we apply Algorithm 1 to obtain the maximum relative difference $r^{\max}$ between $\widehat{\upsilon}^\ell$ and $\widetilde{\upsilon}^\ell$ and the TRO parameter $\theta^{\max}$ at which $r^{\max}$ is attained. If $r^{\max} \le \varepsilon$, we terminate and return the bounding functions $(\widehat{\upsilon}^\ell, \widetilde{\upsilon}^\ell)$. Otherwise, if $r^{\max} > \varepsilon$, the newly identified $\theta^{\max}$ does not belong to Θ_ℓ. Thus, we enlarge the set Θ_ℓ with $\theta^{\max}$ and return to step 1b to solve the TRO-CFLP model with the newly identified TRO parameter $\theta^{\max}$.

In Theorem 3, we show that the Spectrum Search Algorithm converges in a finite number of iterations.

Theorem 3. *Algorithm 2 terminates in a finite number of iterations.*

Theorem 3 shows that Algorithm 2 return a pair of lower and upper bounding functions, $\widehat{\upsilon}$ and $\widetilde{\upsilon}$, with a maximum relative difference $r^{\max} \le \varepsilon$ after a finite number of iterations. When $\varepsilon = 0$, Algorithm 2 outputs the exact optimal value function $\widehat{\upsilon} = \widetilde{\upsilon} = \upsilon^\star$ and the spectrum of optimal solutions to the TRO-CFLP problem. Moreover, in this case, steps 1c and 1d are not needed because we can immediately obtain the exact values of the expected recourse $\widehat{\varphi}^{\mathrm{SP}}(\theta)$ and worst-case recourse $\widehat{\varphi}^{\mathrm{DRO}}(\theta)$ by solving the TRO-CFLP problem. However, solving the TRO-CFLP problem exactly could be computationally challenging for large instances. For such instances, one could set $\varepsilon > 0$. In this case, we obtain a spectrum of ε-optimal solutions. Specifically, let us enumerate the set Θ_ℓ upon termination as $\{\theta_m\}_{m=1}^M$ and denote the corresponding solutions as $\{\widehat{\boldsymbol{o}}_m\}_{m=1}^M$. Then, $\widehat{\boldsymbol{o}}_m$ is an ε-optimal solution on the interval $\Theta_m = \big\{\theta \in [0,1] \mid m \in \arg\min_{m' \in [M]} \big\{ \sum_{i\in I} c^{\mathrm{f}}_i(\widehat{\boldsymbol{o}}_{m'})_i + (1-\theta)\widehat{\varphi}^{\mathrm{SP}}_{m'} + \theta\acute{\varphi}^{\mathrm{DRO}}_{m'} \big\}\big\}$ for all $m \in [M]$.

6. Decomposition Algorithms

In this section, we propose a new hybrid (primal-dual) column-and-constraint generation (hC&CG) algorithm for solving the TRO-CFLP model for a given TRO parameter θ. We utilize this algo-

Algorithm 2: The Spectrum Search Algorithm

Initialization: $\Theta_1 = \{0, 1\}$, iteration index $\ell = 1$, relative tolerance $\varepsilon \geq 0$.

1. Solve TRO-CFLP Problem and Obtain TRO Solution(s).

a. If $\ell = 1$, solve the TRO-CFLP problem with $\theta \in \overline{\Theta} = \{0, 1\}$ to within a relative gap ε, and record the best feasible solution $\widehat{\boldsymbol{o}}(\theta)$, a lower bound $\underline{v}(\theta)$ and an upper bound $\overline{v}(\theta)$ of $v^\star(\theta)$ for all $\theta \in \overline{\Theta}$.

b. If $\ell > 1$, solve the TRO-CFLP problem with TRO parameters $\overline{\Theta} = \{\theta^{\max}\}$ to a relative gap of ε, and record the best feasible solution $\widehat{\boldsymbol{o}}(\theta)$, a lower bound $\underline{v}(\theta)$ and an upper bound $\overline{v}(\theta)$ of $v^\star(\theta)$.

c. Compute $\widehat{\varphi}^{\mathrm{SP}}(\theta) = \mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\widehat{\boldsymbol{o}}(\theta), \boldsymbol{d})]$ or all $\theta \in \overline{\Theta}$.

d. If $\ell > 1$ and $\widehat{\boldsymbol{o}}(\theta^{\max}) \in \{\widehat{\boldsymbol{o}}(\theta) \mid \theta \in \Theta_{\ell-1}\}$, set $\widehat{\varphi}^{\mathrm{DRO}}(\theta^{\max})$ as $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\widehat{\boldsymbol{o}}(\theta^{\max}), \boldsymbol{d})]$. Otherwise, if $\widehat{\boldsymbol{o}}(\theta^{\max}) \notin \{\widehat{\boldsymbol{o}}(\theta) \mid \theta \in \Theta_{\ell-1}\}$, obtain an upper bound of $\widehat{\varphi}^{\mathrm{DRO}}(\theta)$, e.g., (8), for all $\theta \in \overline{\Theta}$.

2. Update the Lower Bounding Values.

a. Sort and enumerate the set Θ_ℓ as $\{\theta_m\}_{m=1}^M$ such that its elements satisfy $0 = \theta_1 < \theta_2 < \cdots < \theta_M = 1$.

b. Set $\underline{v}_m = \underline{v}(\theta_m)$ and $\overline{v}_m = \overline{v}(\theta_m)$ for all $m \in [M]$.

c. If $\widehat{\mathbb{P}}_N \in \mathcal{F}$, update $\underline{v}_m = \max_{m' \leq m}\{\underline{v}_{m'}\}$ for all $m \in [M]$.

d. Update the lower bounding values $\{\underline{v}_m\}_{m=1}^M$ using Algorithm 4 (see Appendix B.1).

3. Update the Lower and Upper Bounding Functions.

a. Update the lower bounding function $\widehat{v}^\ell(\theta)$ as the function linearly interpolating $\{(\theta_m, \underline{v}_m)\}_{m=1}^M$.

b. Update the upper bounding function $\widetilde{v}^\ell(\theta)$ as defined in (9).

4. Identify the New TRO Parameter.

a. Obtain $\theta^{\max}$ and $r^{\max}$ using Algorithm 1.

b. If $r^{\max} \leq \varepsilon$, **terminate** and return the spectrum of location decisions $\{\widehat{\boldsymbol{o}}(\theta) \mid \theta \in \Theta_\ell\}$.

c. If $r^{\max} > \varepsilon$ update $\Theta_{\ell+1} \leftarrow \Theta_\ell \cup \{\theta^{\max}\}$. Set $\ell \leftarrow \ell + 1$ and return to **step 1b**.

rithm as a subroutine (steps 1a and 1b) in our Spectrum Search Algorithm. We first present our proposed hC&CG algorithm in Section 6.1 and then discuss its convergence in Section 6.2.

6.1. The hC&CG Algorithm

Recall that the optimization problem defining $Q(\boldsymbol{o}, \boldsymbol{d})$ in constraints (5d) is an LP. By strong duality, we have $Q(\boldsymbol{o}, \boldsymbol{d}) = \sup_{e\in E}\left\{\sum_{i\in I} C_i v_i^e o_i + \sum_{j\in J} w_j^e d_j\right\}$, where $\{(\boldsymbol{w}^e, \boldsymbol{v}^e)\}_{e\in E}$ is the set of extreme points of the dual feasible set $\Pi = \{(\boldsymbol{w}, \boldsymbol{v}) \mid w_j + v_i \leq t_{i,j},\ 0 \leq w_j \leq c_j^{\mathrm{u}},\ v_i \leq 0,\ \forall i \in I,\ j \in J\}$. Accordingly, we can equivalently reformulate constraints (5d) in (5) as

$$\delta \geq \sum_{i\in I} C_i v_i^e o_i + \sum_{j\in J} w_j^e d_j - \sum_{k\in[K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'} g_{k'}(\boldsymbol{d}), \quad \forall \boldsymbol{d} \in \mathcal{S},\ e \in E. \tag{11}$$

The idea of our hC&CG is to solve a series of relaxations of (5) with a finite but increasing number of constraints corresponding to (11). In particular, in each iteration of the algorithm, we construct two sets of optimality cuts and add them to the current relaxation, thus improving its tightness. Algorithm 3 summarizes the steps of the algorithm. We initialize the algorithm with empty sets $\overline{\mathcal{S}}$ and $\overline{E}$. At iteration h, starting with subsets $\overline{\mathcal{S}} \subseteq \mathcal{S}$ and $\overline{E} \subseteq E$, in step 1.1, we solve the master problem (12) to within a relative gap tolerance of ε_{MP}^h and record the best feasible solution

$(\boldsymbol{o}^h, \boldsymbol{x}^h, \boldsymbol{u}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h)$. Note that the master problem involves two sets of valid optimality cuts. The first set (12b), referred to as *primal optimality cuts*, corresponds to a decomposition over the scenario set $\mathcal{S}$ with the primal form of the second stage problem (see Proposition 5). The second set (12c), referred to as *dual optimality cuts*, corresponds to a decomposition over the dual set E with the dual form of the second-stage problem (see Proposition 5). These cuts are generally non-linear, as they involve maximization problems over $e \in E$ and $\boldsymbol{d} \in \mathcal{S}$. In Proposition 5, we derive equivalent reformulations of these cuts as sets of constraints.

Proposition 5. *For a given $\boldsymbol{d} \in \mathcal{S}$, the primal C&CG optimality cut (12d) is equivalent to*

$$\delta \geq \sum_{i\in I}\sum_{j\in J} t_{i,j}x_{i,j}(\boldsymbol{d}) + \sum_{j\in J} c^{\mathrm{u}}_j u_j(\boldsymbol{d}) - \sum_{k\in[K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'} g_{k'}(\boldsymbol{d}), \tag{14a}$$

$$\sum_{j\in J} x_{i,j}(\boldsymbol{d}) \leq C_i o_i, \quad \sum_{i\in I} x_{i,j}(\boldsymbol{d}) + u_j(\boldsymbol{d}) \geq d_j, \quad \forall i \in I,\, j \in J, \tag{14b}$$

$$x_{i,j}(\boldsymbol{d}) \geq 0, \quad u_j(\boldsymbol{d}) \geq 0, \quad \forall i \in I,\, j \in J. \tag{14c}$$

For a given $e \in E$, the dual C&CG optimality cut (12e) is equivalent to

$$\delta \geq \sum_{i\in I} C_i v_i^e o_i + \sum_{k\in[K]} \rho_k f_k^*(\boldsymbol{\kappa}_k^e) + \sum_{k'\in[K']} \gamma_{k'} g_{k'}^*(\boldsymbol{\tau}_{k'}^e) + \sum_{j\in J} (\overline{\pi}_j \bar{d}_j - \underline{\pi}_j \underline{d}_j), \tag{15a}$$

$$\boldsymbol{w} - \sum_{k\in[K]} \rho_k \boldsymbol{\kappa}_k^e - \sum_{k'\in[K']} \gamma_{k'} \boldsymbol{\tau}_{k'}^e = \overline{\boldsymbol{\pi}} - \underline{\boldsymbol{\pi}}, \tag{15b}$$

$$\boldsymbol{\kappa}_k^e \in \mathbb{R}^{|J|}, \quad \boldsymbol{\tau}_{k'}^e \in \mathbb{R}^{|J|}, \quad \forall k \in [K],\, k' \in [K'], \tag{15c}$$

where f_k^* and $g_{k'}^*$ are the convex conjugates of f_k and $g_{k'}$, respectively.

Remark 1. Constraints (15a) involves two non-linear products $\rho_k f_k^*(\boldsymbol{\kappa}_k^e)$ and $\gamma_{k'} g_{k'}^*(\boldsymbol{\tau}_{k'}^e)$. In Appendix C.2, we show that for some popular ambiguity sets, we can reformulate the optimality cut (12e) and thus, (15), into a set of linear constraints.

Next, after solving the master problem (12), we record a lower bound L^h and upper bound U^h on the optimal value of the master problem (12) from the solver, which will be used in the backtracking routine (step 3) to ensure convergence. If $L^h > \overline{L}$, which certifies that L^h is a valid lower bound on the optimal value $\upsilon^\star(\theta)$ of the TRO-CFLP model, then we set p as the current iteration h. In other words, L^p is always a valid lower bound on $\upsilon^\star(\theta)$. Lastly, we update $\overline{L}$ to U^h, which may accelerate the lower bound improvement in the next master problem; see discussions in Tsang et al. (2023).

In step 2, given the feasible solution $(\boldsymbol{o}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h)$, we solve the subproblem (13) and record an optimal solution $\bar{\boldsymbol{d}} \in \mathcal{S}$, the optimal value $\bar{\delta}$, and an optimal extreme point $(\boldsymbol{w}^{\bar{e}}, \boldsymbol{u}^{\bar{e}}) \in \Pi$; Proposition 6 below shows that such an extreme point always exists. Note that the objective function of (13) is non-convex in general due to the bilinear term $w_j d_j$ for $j \in J$. In Proposition 7, we derive a mixed-integer non-linear programming (MINLP) reformulation of the subproblem (13)

Algorithm 3: The hC&CG algorithm for solving the TRO-CFLP model for a given $\theta \in [0,1]$.

Initialization: $\overline{L} \leftarrow 0$, $\overline{U} \leftarrow \infty$, $\varepsilon > 0$, $\widetilde{\varepsilon} \in (0, \varepsilon/(1+\varepsilon))$, $\{\varepsilon_{MP}^h \in (0,1)\}_{h\in\mathbb{N}}$, $\alpha \in (0,1)$, $\overline{\mathcal{S}} \leftarrow \emptyset$, $\overline{E} \leftarrow \emptyset$, $h \leftarrow 1$, $p \leftarrow 0$.

1. Master Problem.

1.1 Solve the following master problem within a relative optimality gap ε_{MP}^h.

$$\min_{\substack{\boldsymbol{o}\in\mathcal{O},\,\boldsymbol{u}\geq 0,\,\boldsymbol{x}\geq 0,\\ \boldsymbol{\rho},\boldsymbol{\gamma}\geq \mathbf{0},\,\delta}} \quad \sum_{i\in I} c_i^{\mathrm{f}} o_i + \frac{1-\theta}{N}\sum_{n=1}^{N}\Big(\sum_{i\in I}\sum_{j\in J} t_{i,j}x_{i,j}^n + \sum_{j\in J} c_j^{\mathrm{u}} u_j^n\Big) + \theta\Big(\sum_{k\in[K]}\mu_k\rho_k + \sum_{k'\in[K']}\sigma_{k'}\gamma_{k'} + \delta\Big) \tag{12a}$$

$$\text{s.t.} \quad \sum_{j\in J} x_{i,j}^n \leq C_i o_i, \quad \forall i \in I,\, n \in [N], \tag{12b}$$

$$\sum_{i\in I} x_{i,j}^n + u_j^n \geq d_j^n, \quad \forall j \in J,\, n \in [N], \tag{12c}$$

$$\delta \geq \max_{e\in E}\Big\{\sum_{i\in I} C_i v_i^e o_i + \sum_{j\in J} w_j^e d_j - \sum_{k\in[K]}\rho_k f_k(\boldsymbol{d}) - \sum_{k'\in[K']}\gamma_{k'} g_{k'}(\boldsymbol{d})\Big\}, \quad \forall \boldsymbol{d} \in \overline{\mathcal{S}}, \tag{12d}$$

$$\delta \geq \max_{\boldsymbol{d}\in\mathcal{S}}\Big\{\sum_{i\in I} C_i v_i^e o_i + \sum_{j\in J} w_j^e d_j - \sum_{k\in[K]}\rho_k f_k(\boldsymbol{d}) - \sum_{k'\in[K']}\gamma_{k'} g_{k'}(\boldsymbol{d})\Big\}, \quad \forall e \in \overline{E}, \tag{12e}$$

$$\sum_{i\in I} c_i^{\mathrm{f}} o_i + \frac{1-\theta}{N}\sum_{n=1}^{N}\Big(\sum_{i\in I}\sum_{j\in J} t_{ij}x_{i,j}^n + \sum_{j\in J} c_j^{\mathrm{u}} u_j^n\Big) + \theta\Big(\sum_{k\in[K]}\mu_k\rho_k + \sum_{k'\in[K']}\sigma_{k'}\gamma_{k'} + \delta\Big) \geq \overline{L}. \tag{12f}$$

1.2 Record the best feasible solution $(\boldsymbol{o}^h, \boldsymbol{x}^h, \boldsymbol{u}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h, \delta^h)$ found, a lower bound $L^h \geq \overline{L}$, and an upper bound U^h of optimal value of the master problem. If $L^h > \overline{L}$, then set $p \leftarrow h$. Set $\overline{L} \leftarrow U^h$.

2. Subproblem. Solve the following subproblem with $(\boldsymbol{o}^h, \boldsymbol{u}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h)$ obtained from Step 1

$$\sup_{\boldsymbol{d}\in\mathcal{S},\,(\boldsymbol{v},\boldsymbol{w})\in\Pi}\Big\{\sum_{i\in I} C_i v_i o_i^h + \sum_{j\in J} w_j d_j - \sum_{k\in K}\rho_k^h f_k(\boldsymbol{d}) - \sum_{k'\in[K']}\gamma_{k'}^h g_{k'}(\boldsymbol{d})\Big\}, \tag{13}$$

and record an optimal solution $(\overline{\boldsymbol{d}}, \boldsymbol{w}^{\overline{e}}, \boldsymbol{u}^{\overline{e}})$ and optimal value $\overline{\delta}$. Set

$$\overline{U} \leftarrow \min\Big\{\overline{U}, \sum_{i\in I} c_i^{\mathrm{f}} o_i^h + \frac{1-\theta}{N}\sum_{n=1}^{N}\Big(\sum_{i\in I}\sum_{j\in J} t_{i,j}x_{i,j}^{h,n} + \sum_{j\in J} c_j^{\mathrm{u}} u_j^{h,n}\Big) + \theta\Big(\sum_{k\in[K]}\mu_k\rho_k^h + \sum_{k'\in[K']}\sigma_{k'}\gamma_{k'}^h + \overline{\delta}\Big)\Big\}.$$

3. Optimality Check and Backtracking Routine.

If $(\overline{U} - L^p)/\overline{U} < \varepsilon$, terminate. Return the best feasible solution with the best objective value $\overline{U}$. Otherwise, proceed to the exploitation or exploration step below.

- **Exploitation:** If $(\overline{U} - U^h)/\overline{U} < \widetilde{\varepsilon}$, then set $h \leftarrow p$ and $\overline{L} \leftarrow L^p$. Set $\varepsilon_{MP}^h \leftarrow \alpha\varepsilon_{MP}^h$ for all $h \geq p$ and return to **Step 1**.
- **Exploration:** If $(\overline{U} - U^h)/\overline{U} \geq \widetilde{\varepsilon}$, enlarge the scenario set as follows: $\overline{\mathcal{S}} \leftarrow \{\overline{\boldsymbol{d}}\}$ and $\overline{E} \leftarrow \{\overline{e}\}$. Return to **Step 1**.

with convex objective and constraints that can be solved using commercial solvers. Since $(\boldsymbol{o}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h)$ is a feasible solution to (5), its objective value $L^h - \theta(\delta^h - \overline{\delta})$ is a valid upper bound on the optimal value $\upsilon^\star(\theta)$. The best upper bound on $\upsilon^\star(\theta)$ is recorded in $\overline{U}$.

Proposition 6. *Let* $(\overline{\boldsymbol{d}}, \overline{\boldsymbol{w}}, \overline{\boldsymbol{v}}) \in \mathcal{S} \times \Pi$ *be an optimal solution to the subproblem* (13) *obtained using a given feasible solution* $(\boldsymbol{o}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h)$ *to the master problem* (12). *There exists an extreme point* $(\boldsymbol{w}^e, \boldsymbol{v}^e)$ *of* Π *such that* $(\overline{\boldsymbol{d}}, \boldsymbol{w}^e, \boldsymbol{v}^e) \in \mathcal{S} \times \Pi$ *is another optimal solution to* (13) *with* $(\boldsymbol{w}^e, \boldsymbol{v}^e) \in \arg\max_{(\boldsymbol{w},\boldsymbol{v})\in\Pi} \big\{ \sum_{i\in I} C_i v_i o_i^h + \sum_{j\in J} w_j \overline{d}_j \big\}$.

Proposition 7. *Subproblem* (13) *is equivalent to the following MINLP:*

$$\underset{\boldsymbol{u},\boldsymbol{x},\boldsymbol{w},\boldsymbol{v},\boldsymbol{s},\boldsymbol{d}}{\text{maximize}} \quad \sum_{i\in I}\sum_{j\in J} t_{i,j}x_{i,j} + \sum_{j\in J} c_j^{\mathrm{u}} u_j - \sum_{k\in K} \rho_k^h f_k(\boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'}^h g_{k'}(\boldsymbol{d}) \tag{16a}$$

$$\text{subject to} \quad \sum_{j\in J} x_{i,j} \le C_i o_i^h, \quad \forall i \in I, \tag{16b}$$

$$\sum_{i\in I} x_{i,j} + u_j = d_j, \quad \forall j \in J, \tag{16c}$$

$$w_j + v_i \le t_{i,j}, \quad \forall i \in I,\ j \in J, \tag{16d}$$

$$w_j \le c_j^{\mathrm{u}}, \quad \forall j \in J, \tag{16e}$$

$$C_i o_i - \sum_{j\in J} x_{i,j} \le M_i^{1,1} s_i^1,\ -v_i \le M_i^{1,2}(1 - s_i^1), \quad \forall i \in I, \tag{16f}$$

$$t_{i,j} - w_j - v_i \le M_{i,j}^{2,1} s_{i,j}^2,\ x_{i,j} \le M_{i,j}^{2,2}(1 - s_{i,j}^2), \quad \forall i \in I,\ j \in J, \tag{16g}$$

$$c_j^{\mathrm{u}} - w_j \le M_j^{3,1} s_j^3,\ u_j \le M_j^{3,2}(1 - s_j^3), \quad \forall j \in J, \tag{16h}$$

$$x_{i,j} \ge 0,\ u_j \ge 0,\ v_i \le 0,\ s_j^1 \in \{0,1\},\ s_{i,j}^2 \in \{0,1\},\ s_j^3 \in \{0,1\}, \quad \forall i \in I,\ j \in J, \tag{16i}$$

$$\underline{d}_j \le d_j \le \overline{d}_j, \quad \forall j \in J. \tag{16j}$$

Here, the big-M parameters can be chosen as: $M_i^{1,1} = C_i$, $M_i^{1,2} = \max_{j\in J}\{c_j^{\mathrm{u}}\} - \min_{j\in J}\{t_{i,j}\}$, $M_{i,j}^{2,1} = t_{i,j} + M_i^{1,2}$, $M_{i,j}^{2,2} = C_i$, $M_j^{3,1} = c_j^{\mathrm{u}}$, $M_j^{3,2} = \overline{d}_j$ *for all* $i \in I$ *and* $j \in J$.

Finally, in step 3, we first perform an optimality check. Specifically, if the actual relative gap $(\overline{U} - L^p)/\overline{U}$, which is computed based on a valid lower bound L^p and upper bound $\overline{U}$ on $\upsilon^\star(\theta)$, is smaller than the pre-specified tolerance ε, we terminate and return the best feasible solution and the best objective value $\overline{U}$. Otherwise, we proceed to the backtracking routine, which aims to balance the computational gains and inaccuracies from solving the master problems inexactly. Depending on the inexact relative gap $(\overline{U} - U^h)/\overline{U}$, the algorithm proceeds to the exploitation or exploration step. If the inexact relative gap is smaller than the pre-specified tolerance $\widetilde{\varepsilon}$, we exploit the knowledge of the best valid lower bound L^p and return to step 1 by solving the master problem with $\overline{L} = L^p$ and a reduced relative gap tolerance ε_{MP}^h. Otherwise, we enlarge the scenario sets $\overline{\mathcal{S}}$ and $\overline{E}$ with the newly obtained scenarios $\overline{\boldsymbol{d}}$ and $\overline{e}$, respectively, and return to step 1 by solving the master problem with the enlarged scenario sets $\overline{\mathcal{S}}$ and $\overline{E}$.

We close this section by emphasizing that our hC&CG algorithm extends and generalizes existing C&CG algorithms. Specifically, suppose we remove the dual optimality cuts (12e). In this case, our hC&CG algorithm reduces to the primal C&CG (pC&CG) algorithm (see, e.g., Tsang et al.,

2023; Zeng and Zhao, 2013), in which a new set of primal constraints and recourse variables is introduced at each iteration; see the reformulation (14) in Proposition 5. Conversely, if we remove the primal optimality cuts (12d), our hC&CG algorithm reduces to the dual C&CG (dC&CG) algorithm (see, e.g., Saif and Delage, 2021; Tan et al., 2024), where the cut is derived from the dual of the recourse function and involves the convex conjugates of f_k and $g_{k'}$ with the corresponding dual variables $\boldsymbol{\kappa}_k$ and $\boldsymbol{\tau}_{k'}$; see the reformulation (15) in Proposition 5. Benders decomposition (BD) is another classical approach to solving two-stage DRO problems (see, e.g., Jiang et al., 2012; Thiele et al., 2009). The master problem of the BD algorithm incorporates the following optimality cuts

$$\delta \geq \sum_{i\in I} C_i v_i^e o_i + \sum_{j\in J} w_j^e d_j - \sum_{k\in[K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'} g_{k'}(\boldsymbol{d}), \quad \forall \boldsymbol{d}\in\overline{\mathcal{S}},\, e\in\overline{E} \tag{17}$$

instead of optimality cuts (12d)–(12e) in the master problem (12). It is easy to verify that optimality cuts (12d) and (12e) are stronger (i.e., provide better lower bounds) than (17). Indeed, the right-hand sides of (12d) and (12e) are larger than or equal to that of (17). In Section 7, we demonstrate the superior computational performance of our proposed hC&CG algorithm over dC&CG, pC&CG, and BD algorithms (see, e.g., Cheng et al., 2021; Saif and Delage, 2021; Shehadeh and Tucker, 2022).

6.2. Convergence Analysis

In this section, we analyze the convergence of the hC&CG algorithm (i.e., Algorithm 3). In particular, we show that our hC&CG algorithm, as well as the pC&CG, dC&CG, and BD algorithms, converge in a finite number of iterations when the set of feasible solutions $(\boldsymbol{o},\boldsymbol{\rho},\boldsymbol{\gamma},\delta)$ to the master problem is compact. Note that the feasible region $\mathcal{O}$ of variables $\boldsymbol{o}$ is compact, while that of the dual variables $(\boldsymbol{\rho},\boldsymbol{\gamma})$ in (5) are unbounded. In Lemma 1, we show that, without loss of optimality, one can restrict $(\boldsymbol{\rho},\boldsymbol{\gamma})$ to a compact set under the classical Slater-type condition (see, e.g., Xu et al., 2018).

Lemma 1. *Let $\mathcal{P}^+$ be the positive linear space of signed measures generated by $\mathcal{P}(\mathcal{S})$. Suppose that the following Slater-type condition holds:*

$$(1,\mathbf{0}) \in \operatorname{int}\left\{ \Big(\mathbb{E}_{\mathbb{P}}[1], \mathbb{E}_{\mathbb{P}}[\boldsymbol{f}(\boldsymbol{d})], \mathbb{E}_{\mathbb{P}}[\boldsymbol{g}(\boldsymbol{d})]\Big) + \{0\}\times\{0\}^K \times \mathbb{R}_+^{K'} \;\Big|\; \mathbb{P}\in\mathcal{P}^+ \right\}, \tag{18}$$

where $\boldsymbol{f}(\boldsymbol{d}) = \big(f_1(\boldsymbol{d}),\ldots,f_K(\boldsymbol{d})\big)$ and $\boldsymbol{g}(\boldsymbol{d}) = \big(g_1(\boldsymbol{d}),\ldots,g_{K'}(\boldsymbol{d})\big)$. Then, the set of optimal dual solutions to the moment problem $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})]$ is uniformly bounded. That is, there exists $r>0$ (independent of $\boldsymbol{o}$) such that $\mathcal{L}^\star(\boldsymbol{o})\cap r\mathbb{B}(\mathbf{0},1)\neq\emptyset$ for all $\boldsymbol{o}\in\mathcal{O}$, where $\mathcal{L}^\star(\boldsymbol{o})$ is the set of dual optimal solutions $(\boldsymbol{\rho},\boldsymbol{\gamma},\delta)$ and $\mathbb{B}(\mathbf{0},1)$ is a ball centered at $\mathbf{0}$ with radius 1.

Since $\boldsymbol{o}\in\mathcal{O}$ is finite, Lemma 1 shows that, without loss of optimality, we can impose sufficiently large bounds on $(\boldsymbol{\rho},\boldsymbol{\gamma})$, say $\rho_k\in[-M,M]$ and $\gamma_{k'}\in[0,M]$ for all $k\in[K]$ and $k'\in[K']$. Moreover, since we are minimizing δ in reformulation (5) and in the master problem (12), the optimal δ in these problems are also bounded. Hence, we can also impose sufficiently large bounds on δ, say

$\delta \in [-M', M']$. With the compactness of the set of feasible solutions $(\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \delta)$, we are now ready to show the finite convergence of Algorithm 3 in Theorem 4.

Theorem 4. *Suppose that the Slater-type condition* (18) *holds and the optimal solutions to the master problems* (12) *are contained in a compact set. If* $\varepsilon > 0$ *and* $0 < \widetilde{\varepsilon} < \varepsilon/(1+\varepsilon)$*, then the hC&CG algorithm, as well as the pC&CG, dC&CG, and BD algorithms, terminate in a finite number of iterations.*

We make three remarks in order. First, assumptions in Theorem 4 are imposed to ensure that when $\mathcal{S}$ and E are not finite, there is a finite number of exploration steps. However, when $\mathcal{S}$ or E is finite, these assumptions are not needed to guarantee the finite convergence as demonstrated in the proof of Theorem 4. In particular, for the TRO-CFLP model, the set E is finite because the second-stage problem is always feasible. Thus, the finite convergence of the hC&CG and dC&CG algorithms holds without these assumptions. Second, our convergence results in Theorem 4 generalize those of Tsang et al. (2023). Specifically, recall that if we only include primal C&CG optimality cuts (12d) in the master problem (12), our hC&CG algorithm reduces to the pC&CG algorithm of Tsang et al. (2023) tailored for solving the TRO-CFLP model. Tsang et al. (2023) proved the convergence of the pC&CG algorithm when the set of maximizers over $\boldsymbol{d} \in \mathcal{S}$ in the subproblem (13) is finite. In contrast, Theorem 4 establishes finite convergence without requiring any such restriction on $\mathcal{S}$. Finally, the compactness of the (first-stage) feasible region is a typical assumption in classical convergence analyses of cutting-plane algorithms (Blankenship and Falk, 1976; Gustafson and Kortanek, 1973; Mutapcic and Boyd, 2009) and, as discussed above, holds for the TRO-CFLP.

7. Computational Results

In this section, we present extensive computational results demonstrating the computational efficiency of our proposed algorithms. In addition, we show the practical benefits of adopting our TRO-CFLP approach over traditional SP and DRO approaches for the CFLP. In Section 7.1, we discuss the experimental setup. In Sections 7.2 and 7.3, we analyze the computational performance of the proposed hC&CG algorithm and Spectrum Search Algorithm, respectively. Then, in Section 7.4, we analyze the spectrum of optimal solutions obtained using the Spectrum Search Algorithm and compare the operational performance of these solutions via out-of-sample simulation.

7.1. Experimental Setup

We constructed 12 instances based on parameter settings reported in the literature; see Table 7.1. Each instance is characterized by the number of candidate locations $|I|$ and the number of demand nodes $|J|$. These benchmark instances have been used in the related FL literature; see, e.g., Lei et al. (2014) and Shehadeh and Sanci (2021). We estimate the distance between each pair of nodes in each instance as follows. First, we generate a total of $|I| + |J|$ vertices as uniformly distributed random points on a 100-by-100 plane. Then, for each pair of nodes (i, j), we compute

Table 7.1: CFLP instances (Lei et al., 2014; Shehadeh and Sanci, 2021).

Instance	1	2	3	4	5	6	7	8	9	10	11	12
$\|I\|$	5	10	10	20	15	30	20	40	25	50	50	100
$\|J\|$	10	10	20	20	30	30	40	40	50	50	100	100

$t_{i,j}$ as the Euclidean distance between them.

For each instance, we generate the demand's mean $\boldsymbol{\mu}$, lower bound $\underline{\boldsymbol{d}}$, and upper bound $\overline{\boldsymbol{d}}$ as follows. First, we generate μ_j from a Uniform distribution $U[40, 60]$, and set the standard deviation $\nu_j = 0.5\mu_j$, for all $j \in J$. Second, we generate $N' = 1000$ data $\{d_j\}_{n=1}^{N}$ from a lognormal distribution based on (μ_j, ν_j). We then set $\underline{d}_j$ and $\overline{d}_j$ to the 20%-percentile and 80%-percentile of the N' in-sample data for all $j \in J$ in each instance. Next, we generate the samples $\{\widehat{d}_j^n\}_{n=1}^{N}$ for the optimization step from a lognormal distribution based on (μ_j, ν_j) clipped at $[\underline{d}_j, \overline{d}_j]$ for all $j \in J$ and $n \in [N]$. To mimic settings where decision-makers have limited data, we consider a small sample size $N = 10$ (Tsang and Shehadeh, 2023). For illustrative purposes, we consider the mean-support (MS) and mean-absolute-deviation (MAD) as the ambiguity set $\mathcal{F}$ in the TRO-CFLP model (2). We refer to Appendix C for tailored TRO-CFLP models (e.g., reformulations and valid inequalities) with these ambiguity sets. We use the estimated mean and MAD as the input parameters for these sets.

We consider the following diverse combinations of problem parameters: $c_i^{\mathrm{f}} = c^{\mathrm{f}} \in \{1000, 3000, 5000\}$, $C_i = C \in \{100, 200, 300\}$, and $c_j^{\mathrm{u}} = c^{\mathrm{u}} \in \{150, 200, 250, 300\}$ for all $i \in I$ and $j \in J$. These represent different values of fixed cost of opening facilities, facility capacity, and penalty cost of unsatisfied demand, respectively. The unit penalty cost of unmet demand satisfies $c_j^{\mathrm{u}} > \max_{i \in I}\{t_{i,j}\}$ for all $j \in J$. We implement our proposed models and algorithm in MATLAB and use Gurobi (version 11.0.1) as the solver with default settings. We perform all experiments on an HP workstation running Windows Server 2019 with two 2.10 gigahertz Intel (R) Xeon (R) 4208 CPU processors, each with 8 cores (16 total) and 128 gigabytes of shared RAM.

7.2. Analysis of the Computational Performance of the Decomposition Algorithms

In this section, we compare the computational performance of the proposed hC&CG algorithm (introduced in Section 6) with the classical BD, pC&CG, and dC&CG algorithms. We solve instances 1–12 for each combination of the parameters $(c^{\mathrm{f}}, C, c^{\mathrm{u}})$ using these algorithms. That is, we solve each instance 36 times with each algorithm, each time with one of the combinations of $(c^{\mathrm{f}}, C, c^{\mathrm{u}})$. Recall that BD, pC&CG, dC&CG, and hC&CG decompose constraints (11), a reformulation of the worst-case expectation $\sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o}, \boldsymbol{d})]$ in the objective, in a different way. Therefore, we solve each instance with $\theta = 1$ to better evaluate the efficiency of these algorithms. As in Tsang et al. (2023), we set the initial master problem relative gap ε_{MP}^h to 2% and the inexact relative gap $\widetilde{\varepsilon}$ to 1.5%. At each exploitation step, we shrink ε_{MP}^h by $\alpha = 0.5$.

First, we compare the algorithms' performance when solving small- to medium-sized instances,

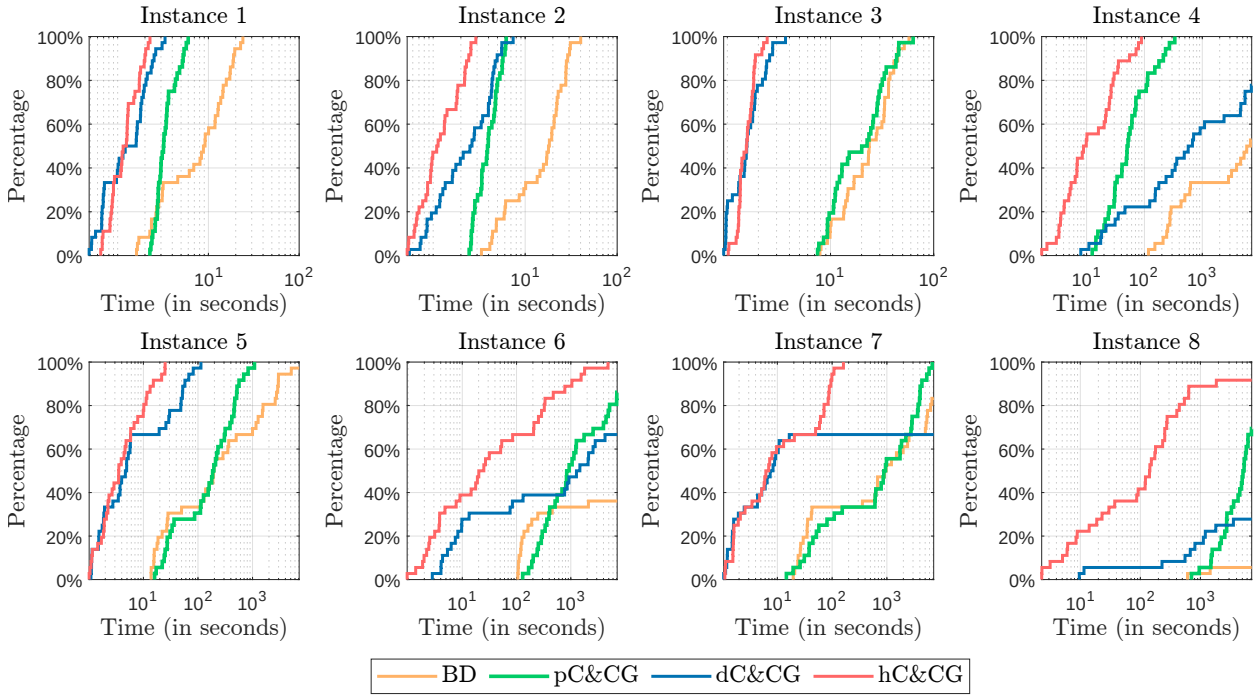


Figure 4: Time performance profile of BD, pC&CG, dC&CG, and hC&CG for instances 1–8. *Note:* The scale of the horizontal axis for instances 1–3 is $[0, 100]$ and for instances 4–8 is $[0, 7200]$.

i.e., instances 1–8. Figure 4 shows the time performance profiles for the four decomposition algorithms when solving these instances with the MS ambiguity set (we observe similar results with the MAD ambiguity set; see Appendix D). The curves represent the percentage of instances solved to a relative gap of $\varepsilon = 2\%$ within a given time $t \in [0, 7200]$. Figure 4 clearly shows that hC&CG has the best computational performance among the four algorithms, while BD has the worst. In particular, BD solution times are significantly longer than the C&CG-based algorithms. Moreover, it can only solve small instances (1–3) and (53, 97, 36, 83, 6)% of instances (4, 5, 6, 7, 8). We attribute BD's inferior performance to having the weakest optimality cut among the algorithms; see Section 6.1. Additionally, we observe that there is no clear winner between pC&CG and dC&CG. For example, while both can solve instances 2, 3, and 5 with all combinations of (C, c^{u}), dC&CG solution times are generally shorter. On the other hand, pC&CG can solve more combinations of instances 4, 6, and 8 than dC&CG within the imposed two-hour time limit.

We also observe that the computational performance of the algorithms varies under different combinations of (C, c^{u}). To see this, in Table 7.2, we present the computational time (in seconds) and the number of iterations required to solve instance 8 with $\varepsilon = 2\%$ and $c^{\mathrm{f}} = 1000$ under different combinations of (C, c^{u}). We do not present results for BD because it failed to solve any of these combinations within two hours. When the capacity C is small ($C = 100$), the computational times of dC&CG are generally shorter than those of pC&CG. However, when the capacity C is larger (i.e., $C \in \{150, 200\}$), dC&CG cannot solve any instance within two hours, with relative gaps ranging

Table 7.2: Computational time (in seconds) and number of iterations when applying pC&CG, dC&CG, and hC&CG to solve instance 8 with $\varepsilon = 2\%$ and $c^{\mathrm{f}} = 1000$ under different combinations of (C, c^{u}). The best computational time among the four algorithms is marked in boldface. *Note:* Solution times with '>' indicate that the instance cannot be solved within two hours.

	$(C, c^{\mathrm{u}}) = (100, 150)$			$(C, c^{\mathrm{u}}) = (150, 150)$			$(C, c^{\mathrm{u}}) = (200, 150)$		
	pC&CG	dC&CG	hC&CG	pC&CG	dC&CG	hC&CG	pC&CG	dC&CG	hC&CG
Time	2099	1120	**30**	5814	>7200	**122**	>7200	>7200	**405**
# Iter	124	133	23	124	226	22	124	644	36
	$(C, c^{\mathrm{u}}) = (100, 200)$			$(C, c^{\mathrm{u}}) = (150, 200)$			$(C, c^{\mathrm{u}}) = (200, 200)$		
	pC&CG	dC&CG	hC&CG	pC&CG	dC&CG	hC&CG	pC&CG	dC&CG	hC&CG
Time	1573	3547	**8**	5373	>7200	**445**	4530	>7200	**92**
# Iter	120	118	12	111	402	34	106	517	24
	$(C, c^{\mathrm{u}}) = (100, 250)$			$(C, c^{\mathrm{u}}) = (150, 250)$			$(C, c^{\mathrm{u}}) = (200, 250)$		
	pC&CG	dC&CG	hC&CG	pC&CG	dC&CG	hC&CG	pC&CG	dC&CG	hC&CG
Time	2224	1805	**19**	4240	>7200	**179**	5337	>7200	**142**
# Iter	132	172	18	104	290	28	114	625	29
	$(C, c^{\mathrm{u}}) = (100, 300)$			$(C, c^{\mathrm{u}}) = (150, 300)$			$(C, c^{\mathrm{u}}) = (200, 300)$		
	pC&CG	dC&CG	hC&CG	pC&CG	dC&CG	hC&CG	pC&CG	dC&CG	hC&CG
Time	3287	555	**9**	2764	>7200	**37**	5116	>7200	**255**
# Iter	157	126	13	112	272	19	108	625	37

from 4.25% to 12.96% upon termination. In contrast, pC&CG can solve most instances within two hours with a significantly fewer number of iterations. These results suggest that the optimality cuts employed in dC&CG are more effective when the capacity C is small, whereas those employed in pC&CG may be stronger when the capacity C is large. These results demonstrate that pC&CG and dC&CG outperform each other under different parameter combinations. Employing both sets of optimality cuts, as in hC&CG, yields the best computational performance. Solution times for hC&CG range from 10 to 50 seconds and are, on average, 85 and 90 times faster than pC&CG and dC&CG. These results demonstrate the computational advantages of our new hC&CG algorithm.

Finally, we investigate the computational performance of the three C&CG algorithms on solving larger instances. Figure 5 shows the time performance profiles of the three C&CG algorithms for solving instances 9–12 (see Appendix D for similar results with the MAD ambiguity set). We do not present results for BD as it failed to solve any of these instances within the imposed two-hour time limit. Clearly, hC&CG has the best computational performance among the three algorithms, while pC&CG has the worst. In particular, hC&CG can solve a significantly larger number of instances than pC&CG and dC&CG. Moreover, hC&CG solution times are considerably shorter. Interestingly, dC&CG can solve more instances at a faster rate than pC&CG. Moreover, pC&CG cannot solve instance 12 (the largest instance) within two hours under any parameter combination.

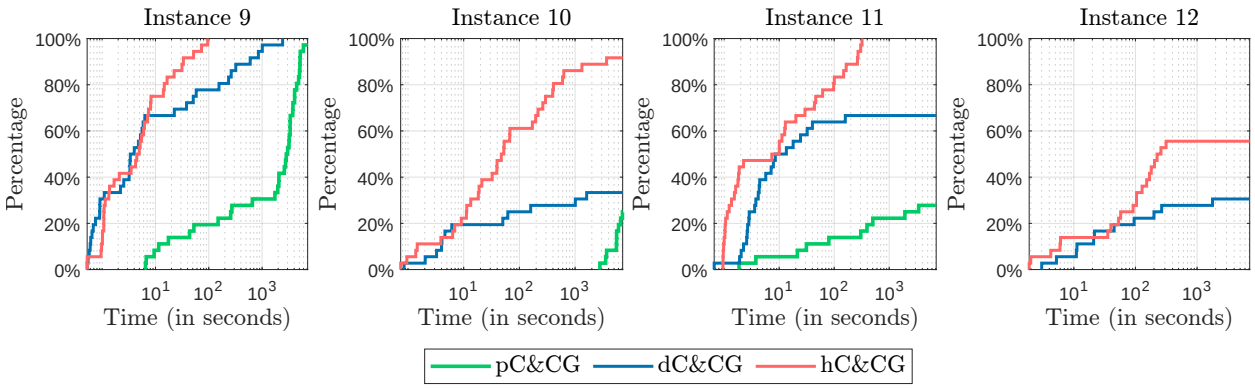


Figure 5: Time performance profile for pC&CG, dC&CG, and hC&CG when solving instances 9–12 with the MS ambiguity set and $\varepsilon = 5\%$

The results in this section demonstrate the superior performance of our newly proposed hC&CG algorithm compared to three state-of-the-art benchmark methods—BD, pC&CG, and dC&CG—which have been widely applied to solve DRO models for various facility location problems (Cheng et al., 2021; Saif and Delage, 2021; Shehadeh and Tucker, 2022).

7.3. Analysis of the Spectrum Search Algorithm

Let us now analyze the computational performance of the Spectrum Search Algorithm (Algorithm 2). We employ hC&CG in step 1 of Algorithm 2 to solve the TRO-CFLP model because, as demonstrated in the previous section, it has the best computational performance among the four decomposition algorithms. First, we demonstrate the algorithm's convergence using instance 4 with $(c^{\mathrm{f}}, C, c^{\mathrm{u}}) = (5000, 100, 300)$, $\varepsilon = 0.01\%$, and a TRO set constructed using the empirical distribution based on data $\{\widehat{\boldsymbol{d}}^n\}_{n=1}^N$ and the MS ambiguity set (see Section 7.1 for details). The algorithm terminates in six iterations with the relative difference between the final lower and upper bounding functions equals 0.01% for any $\theta \in [0, 1]$. Figure 6 shows the improvement of the lower and upper bounding functions computed in step 3 of Algorithm 2. Each plot in the figure corresponds to an iteration and shows the lower and upper bounding functions from the previous iteration (dotted lines) and the current iteration (solid lines). The vertical line indicates the value of $\theta^{\max}$ identified in the previous iteration. It is clear from Figure 6 that in each iteration, the Spectrum Search Algorithm identifies the value of $\theta^{\max}$ where the relative difference between the lower and upper bounding functions is the largest. Then, in the next iteration, the lower and upper bounding functions are improved by solving the TRO-CFLP model with $\theta^{\max}$.

Next, we analyze the computational performance of the Spectrum Search Algorithm. For illustrative purposes, Table 7.3 presents the computational time (in seconds) and the number of TRO parameters θ identified (including 0 and 1) when applying the Spectrum Search Algorithm to solve instances 8 and 9 with $c^{\mathrm{f}} = 5000$ under different combinations of (C, c^{u}) using the MS and MAD ambiguity sets. We observe the following from the results in Table 7.3. First, the Spectrum Search

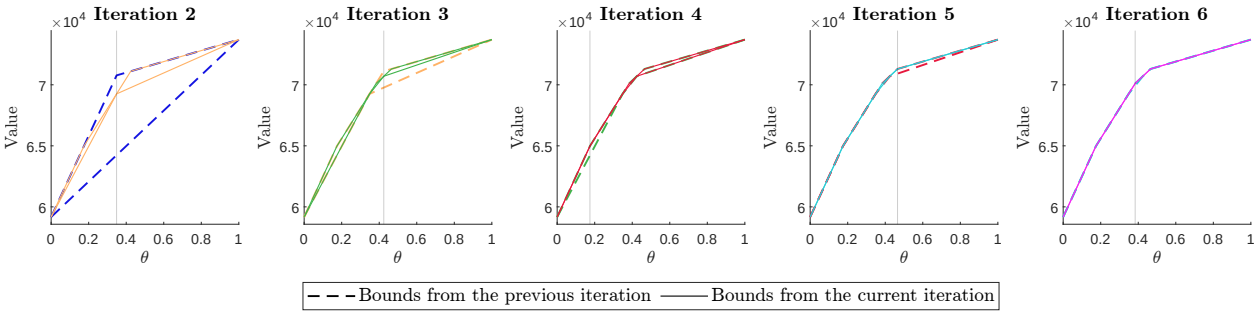


Figure 6: Illustration of the convergence of the Spectrum Search Algorithm for instance 4 with $(c^{\mathrm{f}}, C, c^{\mathrm{u}}) = (5000, 100, 300)$ and $\varepsilon = 0.01\%$ using the MS ambiguity set

Algorithm can efficiently identify the spectrum of optimal solutions to the TRO-CFLP model within a reasonable time. Specifically, the solution times for solving instances 8 and 9 range from 20 to 2764 seconds and from 7 to 384 seconds, respectively. Moreover, we observe that, in general, the solution time for solving the TRO-CFLP model with a larger θ is longer. For example, the solution time for solving instance 8 using the TRO-CFLP model with $(c^{\mathrm{f}}, C, c^{\mathrm{u}}) = (5000, 150, 250)$ and $\theta = (0, 0.300, 0.323, 0.343, 1)$ is $(3, 42, 126, 307, 667)$ seconds. Second, the number of TRO parameters θ identified (and hence the number of TRO-CFLP models solved) ranges from three to eight. This indicates that the number of distinct solutions on the spectrum of optimal solutions for each instance is finite. In particular, each distinct solution on the spectrum corresponds to a different range of θ. For example, the spectrum of optimal solutions for instance 4 with $(c^{\mathrm{f}}, C, c^{\mathrm{u}}) = (5000, 150, 300)$ has four distinct solutions on the intervals $\theta \in [0, 0.175]$, $\theta \in [0.175, 0.384]$, $\theta \in [0.384, 0.466]$, and $\theta \in [0.466, 1]$. Solving the TRO-CFLP model with arbitrarily large numbers of pre-specified values of θ within the same range yields repeated (the same) solutions, i.e., a wasted computational effort in exploring the repeated solutions. One can identify distinct solutions more efficiently using the Spectrum Search Algorithm.

These results motivate the need for our Spectrum Search Algorithm and demonstrate its computational efficiency in identifying the entire spectrum of distinct optimal solutions to the TRO-CFLP model.

7.4. Analysis of the Optimal Spectrum of Solutions

Let us now analyze the spectrum of optimal solutions obtained from solving the TRO-CFLP model using the Spectrum Search Algorithm. We use instance 8 with $(c^{\mathrm{f}}, C, c^{\mathrm{u}}) = (5000, 200, 300)$ for illustrative purposes and brevity. In addition, we use the MS ambiguity set in the TRO set and $\varepsilon = 2\%$ in the algorithm. We observe that the spectrum of optimal solutions has four distinct solutions on the intervals $\theta \in [0, 0.142]$, $\theta \in [0.142, 0.226]$, $\theta \in [0.226, 0.240]$, and $\theta \in [0.240, 1]$. We denote these solutions as Sol-1, Sol-2, Sol-3, and Sol-4, respectively, where Sol-1 is the most optimistic (as all solutions on the interval $\theta \in [0, 0.142]$ are the same as the SAA solution with $\theta = 0$) and Sol-4 is the most conservative. We present these solutions in Figure 7, where the leftmost plot

Table 7.3: Computational time (in seconds) and the number of TRO parameters θ identified (including 0 and 1) when applying the Spectrum Search Algorithm to solve instances 8 and 9 with $\varepsilon = 2\%$ and $c^{\mathrm{f}} = 5000$ under different combinations of (C, c^{u}) using the MS and MAD ambiguity sets. *Note:* Instances marked with "†" are solved to $\varepsilon = 3\%$.

Instance 8		MS				MAD			
$c^{\mathrm{f}} = 5000$		$c^{\mathrm{u}} = 150$	$c^{\mathrm{u}} = 200$	$c^{\mathrm{u}} = 250$	$c^{\mathrm{u}} = 300$	$c^{\mathrm{u}} = 150$	$c^{\mathrm{u}} = 200$	$c^{\mathrm{u}} = 250$	$c^{\mathrm{u}} = 300$
$C = 100$	Time	22	20	26	36	37	94	45	22
	# Identified θ	3	6	8	6	3	8	8	5
$C = 150$	Time	561	861	†1144	845	211	140	176	125
	# Identified θ	6	7	3	5	6	7	6	6
$C = 200$	Time	398	446	186	640	2764	1384	1292	725
	# Identified θ	6	6	5	5	6	7	6	6
Instance 9		MS				MAD			
$c^{\mathrm{f}} = 5000$		$c^{\mathrm{u}} = 150$	$c^{\mathrm{u}} = 200$	$c^{\mathrm{u}} = 250$	$c^{\mathrm{u}} = 300$	$c^{\mathrm{u}} = 150$	$c^{\mathrm{u}} = 200$	$c^{\mathrm{u}} = 250$	$c^{\mathrm{u}} = 300$
$C = 100$	Time	8	7	24	28	11	14	124	101
	# Identified θ	2	3	3	3	2	3	3	3
$C = 150$	Time	38	27	49	109	30	96	84	94
	# Identified θ	5	5	6	7	3	6	7	8
$C = 200$	Time	153	130	185	384	368	214	176	95
	# Identified θ	6	7	8	7	6	7	8	6

shows the SAA solution (Sol-1) and the rightmost plot shows the DRO solution (Sol-4). In each plot, dots represent demand locations, and diamonds represent potential facility locations. Filled diamonds indicate opened facilities, while hollow diamonds indicate locations with no facilities. Except for the leftmost plot, a red circle around a filled diamond indicates that a new facility is opened at this location in this solution compared with the one on the left. A gray circle around a hollow diamond indicates that the previous solution has an open facility at this location.

It is clear that the TRO-CFLP model opens more facilities with a larger θ, i.e., as the model conservativeness increases. Specifically, the number of opened facilities in Sol-1 (leftmost plot), Sol-2, Sol-3, and Sol-4 (rightmost plot) solutions on the spectrum are 10, 11, 12, and 13, respectively. However, these solutions have slight differences in the location of open facilities. For example, compared with Sol-2 (corresponding to $\theta \in [0.142, 0.226]$), which has 11 opened facilities, Sol-3 (corresponding to $\theta \in [0.226, 0.240]$) has one additional facility opened at a new location. Moreover, one facility in Sol-3 are opened near the location with an opened facility in Sol-2 (highlighted with a gray circle). The remaining facilities in Sol-3 are opened at the same locations as in Sol-2.

Next, we investigate the sensitivity of the spectrum of optimal solutions to two key input parameters: the capacity of each facility, C, and the unit penalty cost of unmet demand, c^{u}. Figure 8 presents the optimal number of opened facilities for instance 8 with $c^{\mathrm{f}} = 5000$ under different combinations of (C, c^{u}). We observe the following from this figure. First, the model

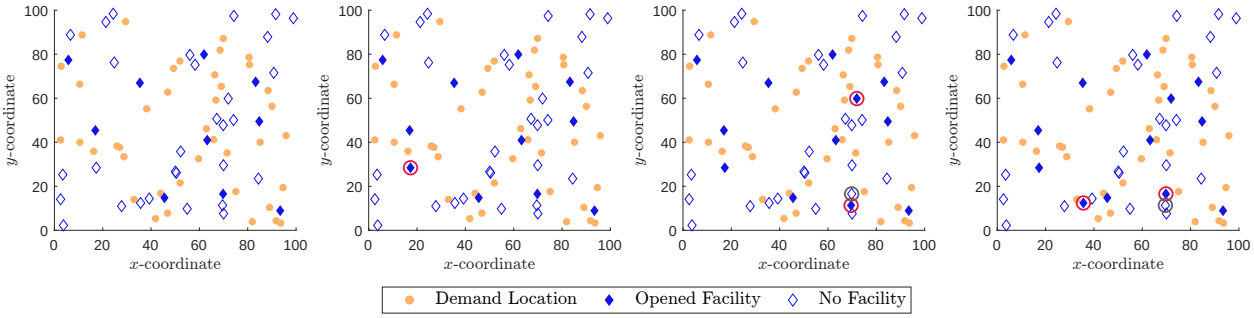


Figure 7: The four distinct solutions on the spectrum of optimal solutions for instance 8 with $(c^{\mathrm{f}}, C, c^{\mathrm{u}}) = (5000, 200, 300)$ using the MS ambiguity set.

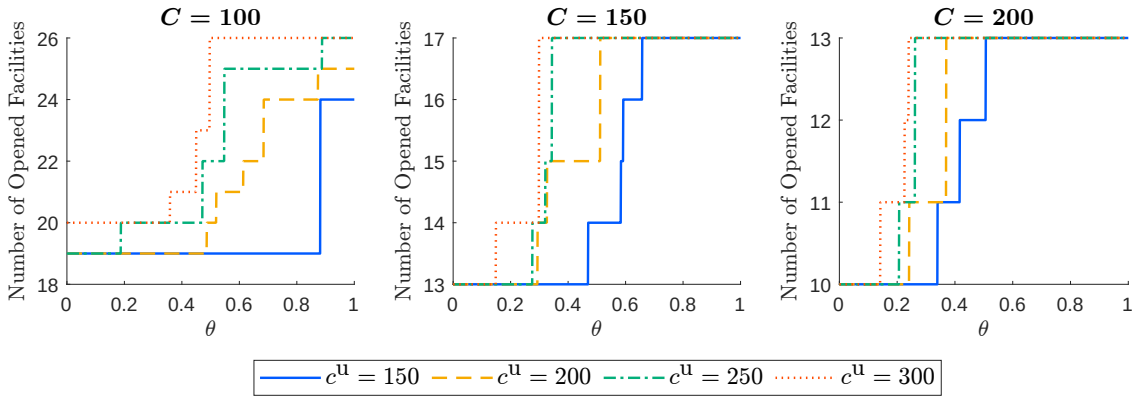


Figure 8: The number of opened facilities for instance 8 with $c^{\mathrm{f}} = 5000$ and $\varepsilon = 2\%$ using the MS ambiguity set under different combinations of (C, c^{u})

produces a different spectrum of solutions with different combinations of (C, c^{u}). Second, for each value of the TRO parameter θ, the model opens more facilities when c^{u} is higher and C is tighter to mitigate the risk of shortage. Third, interestingly, for each combination of (C, c^{u}), there exists a cut-off value of θ at which the model becomes more conservative and opens more facilities than the SAA solution ($\theta = 0$). This cut-off value generally decreases as c^{u} increases. Fourth, when capacity is tight and c^{u} is low, the model yields a spectrum with fewer distinct solutions than when c^{u} is higher. Specifically, when $C = 100$ and $c^{\mathrm{u}} = 150$, the spectrum consists of two distinct solutions, and for $\theta \in [0, 0.882]$, the number of opened facilities is the same as in the SAA solution. In contrast, the spectrum of optimal solutions consists of 6, 5, and 4 distinct solutions when $c^{\mathrm{u}} = 200$, 250, and 300, respectively.

7.5. Analysis of the Out-of-Sample Performance

As mentioned earlier, each distinct solution on the spectrum of optimal solutions to the TRO-CFLP model has different fixed costs and operational performance (second-stage costs) and hence, different total costs. To see this, in this section, we analyze the performance of the spectrum of optimal solutions presented in Figure 7 via out-of-sample simulation testing; we have similar observations for other instances. Specifically, we first generate $N' = 10,000$ out-of-sample scenarios

following the same procedure described in Section 7.1, except that we change the generated mean from μ_j to $\mu_j + \Delta$ for all $j \in J$. A positive (resp. negative) Δ corresponds to an increase (resp. decrease) in mean, which mimics the setting where there is a shift in the demand distributions. We consider $\Delta \in \{-5, -4, \ldots, 20\}$ in this experiment. Then, we solve the second-stage problem with the generated N' scenarios to compute the out-of-sample costs associated with each solution on the spectrum. For brevity and illustrative purposes, in Figure 9, we present the average out-of-sample second-stage cost (travel cost plus the unsatisfied demand penalty cost) and total cost (first-stage cost plus second-stage cost) for the spectrum of solutions presented earlier in Figure 7.

We observe the following from Figure 9. When the demand deviations are extremely large ($\Delta \geq 12$), Sol-4 leads to significantly lower second-stage costs (indicating the best operational performance) and total costs. For slight and negative deviations in demand ($\Delta \leq 1$), Sol-1 (SAA solution) results in slightly higher second-stage costs than the other solutions but still offers slightly lower total costs. This is reasonable because when Δ is close to zero, the in-sample distribution (used to generate data for optimization) and out-of-sample distribution (representing the actual distribution) are similar, and when Δ is negative, the actual demand is lower. In both cases, opening more facilities, as in other solutions, will only increase the fixed cost without significantly improving the operational performance. However, Sol-1 performance significantly deteriorates, leading to higher second-stage and total costs when the actual demand distribution deviates from the in-sample data, even under small positive deviations.

When demand variations are between these extremes, Sol-2 and Sol-3 outperform the others, leading to the lowest total costs. Specifically, when $\Delta \in [2, 6]$, the second-stage costs for Sol-2 are significantly lower than those of Sol-1 and slightly higher than those of Sol-3 and Sol-4. However, the total cost of Sol-2 is lower than those of the other solutions because fewer facilities are opened compared with Sol-3 and Sol-4, resulting in lower fixed costs, while more facilities are opened compared with Sol-1, which leads to lower unmet demand penalties that offset the increase in fixed costs. When $\Delta \in [7, 11]$, the second-stage costs for Sol-3 are significantly lower than those of Sol-1 and Sol-2 and slightly higher than those of Sol-4. However, the total costs of Sol-3 are lower than those of the other solutions. This is again because fewer facilities are opened compared with Sol-4, resulting in lower fixed costs. In comparison, more facilities are opened than Sol-1 and Sol-2, which leads to lower unmet demand penalties that offset the increase in fixed costs.

These results demonstrate the benefits of adopting solutions on the spectrum of TRO-CFLP optimal solutions over those obtained via the traditional SAA and DRO models for the CFLP. In particular, if the actual distribution deviates moderately from the in-sample distribution, which is more likely to occur in practice compared with extreme distributional shifts, the total costs of SAA and DRO solutions (e.g., Sol-1 and Sol-4) could be higher than those of the intermediate solutions (e.g., Sol-2 and Sol-3) on the spectrum.

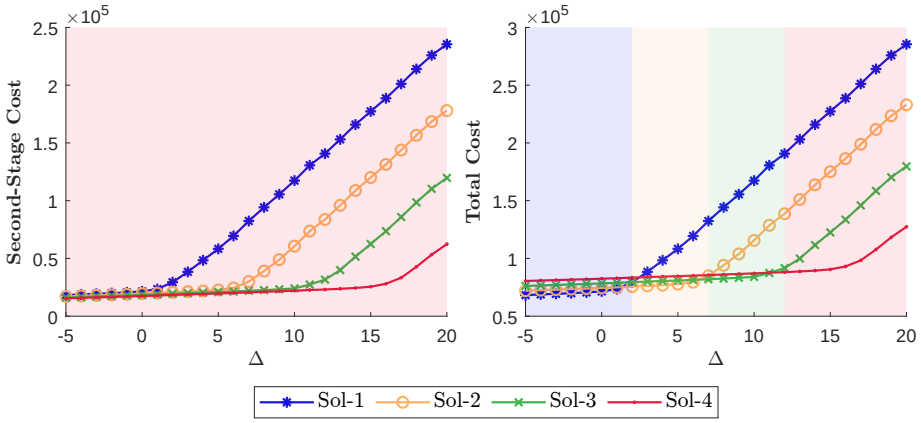


Figure 9: Average out-of-sample second-stage and total costs for instance 8 with $(c^{\mathrm{f}}, C, c^{\mathrm{u}}) = (5000, 200, 300)$ and $\varepsilon = 2\%$ using the MS ambiguity set under different Δ. The background color in the plots indicates the solution with the best performance.

8. Conclusion

In this paper, we introduce innovative algorithmic strategies for identifying the trade-off between adopting a distributional belief and hedging against distributional ambiguity for modeling demand uncertainty in the CFLP. We begin our investigations by introducing a TRO-CFLP model for the CFLP. This model is equipped with a TRO ambiguity set characterized by the empirical distribution, an ambiguity set representing distributional ambiguity, and a TRO parameter that controls the trade-off between solving the problem under the empirical distribution and solving it under the worst-case scenario within the ambiguity. We identify the structural properties of the TRO-CFLP model and propose a Spectrum Search Algorithm that efficiently identifies the entire spectrum of optimal solutions for the model. Additionally, we propose a new hybrid primal-dual column-and-constraint generation (hC&CG) framework for solving the TRO-CFLP model with a pre-specified value of the TRO parameter. This algorithm is employed as a subroutine within the Spectrum Search algorithm. Our extensive computational results demonstrate the computational efficiency of our proposed algorithms in identifying the full spectrum of optimal solutions for the TRO-CFLP model and solving the TRO-CFLP model. Our results also demonstrate the practical benefits of adopting solutions on the TRO-CFLP's spectrum over those from traditional SAA and DRO models. Our proposed algorithms can be customized for other applications.

Our paper presents a new step toward designing computationally efficient decomposition algorithms for identifying the full spectrum of optimal solutions to TRO models and for solving two-stage DRO and TRO problems. We suggest the following areas for future investigations. First, future studies could build on our theoretical results to extend the Spectrum Search Algorithm to address problems involving continuous or mixed-integer first-stage decisions and mixed-integer second-stage recourse. Second, generalizing the hC&CG algorithm to handle two-stage DRO models with integer recourse would offer valuable contributions.

Appendix A. Mathematical Proofs

Appendix A.1. Proof of Theorem 1

Proof. Proof. First, we prove assertion (a). Note that we can write $v^\star(\theta) = \min_{\boldsymbol{o}\in\mathcal{O}}\{f(\boldsymbol{o};\theta)\}$, where

$$\begin{aligned} f(\boldsymbol{o};\theta) &= \sum_{i\in I} c^{\mathrm{f}} o_i + (1-\theta)\mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\boldsymbol{o},\boldsymbol{d})] + \theta \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})] \\ &= \sum_{i\in I} c^{\mathrm{f}} o_i + \mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\boldsymbol{o},\boldsymbol{d})] + \theta\left\{ \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})] - \mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\boldsymbol{o},\boldsymbol{d})] \right\} \end{aligned}$$

is the objective function of the TRO model with TRO parameter θ. It is easy to see that for any $\boldsymbol{o} \in \mathcal{O}$, the function $f(\boldsymbol{o};\theta)$ is linear in θ. Since the feasible set $\mathcal{O} \subseteq \{0,1\}^{|I|}$ is finite, $v^\star(\theta)$ is the pointwise minimum of a finite number of linear functions. It follows that $v^\star$ is piecewise linear and concave. Next, we prove assertion (b). Since $\widehat{\mathbb{P}}_N \in \mathcal{F}$, we have $\mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\boldsymbol{o},\boldsymbol{d})] \leq \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})]$. Therefore, $f(\boldsymbol{o};\theta)$ is non-decreasing in θ for any $\boldsymbol{o} \in \mathcal{O}$. This shows that $v^\star$ is non-decreasing and completes the proof. □

Appendix A.2. Proof of Proposition 1

Proof. Proof. Note that we can rewrite the TRO-CFLP model with the TRO set $\mathcal{F}'(\theta)$ constructed using the moment-based ambiguity set $\mathcal{F}$ defined in (4) as

$$v^\star(\theta) = \underset{\boldsymbol{o}\in\mathcal{O}}{\text{minimize}} \left\{ \sum_{i\in I} c_i^{\mathrm{f}} o_i + (1-\theta)\cdot \mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\boldsymbol{o},\boldsymbol{d})] + \theta \cdot \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})] \right\}. \tag{A.1}$$

Reformulating the second term $\mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\boldsymbol{o},\boldsymbol{d})]$ in (A.1) is straightforward. Thus, we focus on the reformulation of the third term $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})]$, which is equivalent to

$$\begin{aligned} \underset{\mathbb{P}\in\mathcal{P}^+}{\text{maximize}} \quad & \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})] && \text{(A.2a)} \\ \text{subject to} \quad & \mathbb{E}_{\mathbb{P}}[1] = 1, && \text{(A.2b)} \\ & \mathbb{E}_{\mathbb{P}}[f_k(\boldsymbol{d})] = \mu_k, \quad \forall k \in [K], && \text{(A.2c)} \\ & \mathbb{E}_{\mathbb{P}}[g_{k'}(\boldsymbol{d})] \leq \sigma_{k'}, \quad \forall k' \in [K'], && \text{(A.2d)} \end{aligned}$$

where $\mathcal{P}^+$ is the positive linear space of signed measures generated by $\mathcal{P}(\mathcal{S})$. Let δ, ρ_k, and $\gamma_{k'}$ be the dual variables associated with constraints (A.2b), (A.2c), and(A.2d), respectively. Since the set $\mathcal{S}$ is compact and the functions $\{f_k\}_{k\in[K]}$ and $\{g_{k'}\}_{k'\in[K']}$ are continuous, by strong duality of conic linear problems (Shapiro, 2001), (A.2) is equivalent to its dual problem:

$$\begin{aligned} \underset{\boldsymbol{\rho},\boldsymbol{\gamma}\geq\mathbf{0},\,\delta}{\text{minimize}} \quad & \boldsymbol{\mu}^\top\boldsymbol{\rho} + \boldsymbol{\sigma}^\top\boldsymbol{\gamma} + \delta && \text{(A.3a)} \\ \text{subject to} \quad & \delta \geq Q(\boldsymbol{o},\boldsymbol{d}) - \sum_{k\in[K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'} g_{k'}(\boldsymbol{d}), \quad \forall \boldsymbol{d} \in \mathcal{S}. && \text{(A.3b)} \end{aligned}$$

Replacing $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})]$ with its equivalent reformulation (A.3) and $\mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\boldsymbol{o},\boldsymbol{d})]$ with its SAA formulation, we show that (A.1) is equivalent to (5). □

Appendix A.3. Proof of Proposition 2

Proof. Proof. We first prove part (a). We start by proving that $\widehat{\upsilon}$ is the function linearly interpolating the extreme points of $\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)$. Note that $\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)$ is closed and $(0,-1)$ is the only extreme direction of $\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)$. It follows from Theorem 5.36 of Güler (2010) that the closed convex set $\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)$ is the (Minkowski) sum of the convex hull of its extreme points and extreme directions, i.e.,

$$\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big) = \operatorname{conv}\Big(\mathcal{E}\big(\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)\big)\Big) + \{t\cdot(0,-1) \mid t\geq 0\},$$

where $\mathcal{E}\big(\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)\big)$ is the set of extreme points of $\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)$. By the definition of $\widehat{\upsilon}$ in (7a), we have

$$\begin{aligned}
\widehat{\upsilon}(\theta) &= \sup\big\{z \;\big|\; (\theta,z)\in \operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)\big\}\\
&= \sup\Big\{z \;\Big|\; (\theta,z)\in \operatorname{conv}\Big(\mathcal{E}\big(\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)\big)\Big) + \{t\cdot(0,-1)\mid t\geq 0\}\Big\}\\
&= \sup\Big\{z \;\Big|\; (\theta,z)\in \operatorname{conv}\Big(\mathcal{E}\big(\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)\big)\Big)\Big\}.
\end{aligned}$$

Since the number of extreme points of $\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)$ is finite, this implies that $\widehat{\upsilon}$ is the function linearly interpolating the extreme points of $\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)$. In particular, $\widehat{\upsilon}$ is piecewise linear and concave.

Next, we show that $\widehat{\upsilon}\leq \upsilon^\star$. Let $\{(\theta_{m_r},\underline{\upsilon}_{m_r}\}_{r=1}^R$ be the set of extreme points of $\operatorname{conv}\big(\operatorname{hyp}_{[0,1]}(\breve{\upsilon})\big)$ with $0=\theta_{m_1}<\theta_{m_2}<\cdots<\theta_{m_R}=1$. Consider the interval $(\theta_{m_r},\theta_{m_{r+1}})$ for any $r\in[R-1]$. Note that $\widehat{\upsilon}$ is a line joining the extreme points $(\theta_{m_r},\underline{\upsilon}_{m_r})$ and $(\theta_{m_{r+1}},\underline{\upsilon}_{m_{r+1}})$. Suppose, on the contrary, that there exists $\theta'\in(\theta_{m_r},\theta_{m_{r+1}})$ such that $\upsilon^\star(\theta')<\widehat{\upsilon}(\theta')$. Let us write $\theta'=(1-\alpha)\theta_{m_r}+\alpha\theta_{m_{r+1}}$ with $\alpha=(\theta'-\theta_{m_r})/(\theta_{m_{r+1}}-\theta_{m_r})\in(0,1)$. Then, we have

$$\begin{aligned}
\upsilon^\star(\theta')<\widehat{\upsilon}(\theta') &= \frac{\theta_{m_{r+1}}-\theta'}{\theta_{m_{r+1}}-\theta_{m_r}}\,\underline{\upsilon}_{m_r} + \frac{\theta'-\theta_{m_r}}{\theta_{m_{r+1}}-\theta_{m_r}}\,\underline{\upsilon}_{m_{r+1}}\\
&\leq \frac{\theta_{m_{r+1}}-\theta'}{\theta_{m_{r+1}}-\theta_{m_r}}\,\upsilon^\star(\theta_{m_r}) + \frac{\theta'-\theta_{m_r}}{\theta_{m_{r+1}}-\theta_{m_r}}\,\upsilon^\star(\theta_{m_{r+1}})\\
&= (1-\alpha)\upsilon^\star(\theta_{m_r}) + \alpha\upsilon^\star(\theta_{m_{r+1}})
\end{aligned}$$

which contradicts the concavity of $\upsilon^\star$ in Theorem 1. This completes the proof of part (a).

Now, we prove part (b). Since $\widehat{\mathbb{P}}_N\in\mathcal{F}$, if follows from Theorem 1 that $\upsilon^\star$ is non-decreasing. Therefore, for any fixed $m\in[M]$, we have $\underline{\upsilon}_{m'}\leq\upsilon^\star(\theta_{m'})\leq\upsilon^\star(\theta_m)$ for all $m'\leq m$. This shows that $\underline{\upsilon}'_m=\max_{m'\leq m}\underline{\upsilon}_{m'}$ is a valid lower bound on $\upsilon^\star(\theta_m)$. It follows from part (a) that the function $\widehat{\upsilon}(\theta)$ defined in (7a) constructed based on $\{(\theta_m,\underline{\upsilon}'_m)\}_{m=1}^M$ is piecewise linear and concave with $\widehat{\upsilon}^\varepsilon\leq\upsilon^\star$. Moreover, since $\{\underline{\upsilon}'_m\}_{m=1}^M$ is non-decreasing by construction, $\widehat{\upsilon}(\theta)$ is also non-decreasing. □

Appendix A.4. Proof of Proposition 3

Proof. Proof. First, we prove part (a). By definition, $\widetilde{v}(\theta)$ is the pointwise minimum of the linear functions $\widehat{f}(\widehat{\boldsymbol{o}}_m;\theta) = \sum_{i\in I} c^{\mathrm{f}}(\widehat{\boldsymbol{o}}_m)_i + (1-\theta)\widehat{\varphi}_m^{\mathrm{SP}} + \theta\acute{\varphi}_m^{\mathrm{DRO}}$ for $m \in [M]$. This shows that $\widetilde{v}$ is piecewise linear and concave. From the definition of $\widetilde{v}(\theta)$, we have

$$\begin{aligned}
\widetilde{v}(\theta) &= \min_{m\in[M]} \left\{ \sum_{i\in I} c^{\mathrm{f}}(\widehat{\boldsymbol{o}}_m)_i + (1-\theta)\widehat{\varphi}_m^{\mathrm{SP}} + \theta\acute{\varphi}_m^{\mathrm{DRO}} \right\} \\
&\geq \min_{m\in[M]} \left\{ \sum_{i\in I} c^{\mathrm{f}}(\widehat{\boldsymbol{o}}_m)_i + (1-\theta)\widehat{\varphi}_m^{\mathrm{SP}} + \theta\widehat{\varphi}_m^{\mathrm{DRO}} \right\} \\
&= \min_{m\in[M]} \left\{ \sum_{i\in I} c^{\mathrm{f}}(\widehat{\boldsymbol{o}}_m)_i + (1-\theta)\mathbb{E}_{\widehat{\mathbb{P}}_N}\mathbb{E}[Q(\widehat{\boldsymbol{o}}_m,\boldsymbol{d})] + \theta \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\widehat{\boldsymbol{o}}_m,\boldsymbol{d})] \right\} \\
&\geq \min_{\boldsymbol{o}\in\mathcal{O}} \left\{ \sum_{i\in I} c^{\mathrm{f}} o_i + (1-\theta)\mathbb{E}_{\widehat{\mathbb{P}}_N}\mathbb{E}[Q(\boldsymbol{o},\boldsymbol{d})] + \theta \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})] \right\} = v^\star(\theta),
\end{aligned}$$

where the first inequality follows from $\acute{\varphi}_m^{\mathrm{DRO}} \geq \widehat{\varphi}_m^{\mathrm{DRO}}$, and the last inequality follows from the fact that $\{\boldsymbol{o}_m^\star\}_{m=1}^M$ is a subset of $\mathcal{O}$.

Next, we prove part (b). Since $\widehat{\mathbb{P}}_N \in \mathcal{F}$, we have $\acute{\varphi}_m^{\mathrm{DRO}} - \widehat{\varphi}_m^{\mathrm{SP}} \geq \widehat{\varphi}_m^{\mathrm{DRO}} - \widehat{\varphi}_m^{\mathrm{SP}} = \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\widehat{\boldsymbol{o}}_m,\boldsymbol{d})] - \mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\widehat{\boldsymbol{o}}_m,\boldsymbol{d})] \geq 0$ for all $m \in [M]$. Thus, $\widetilde{v}$ is the pointwise minimum of non-decreasing linear functions and hence is non-decreasing. □

Appendix A.5. Proof of Proposition 4

Proof. Proof. We leverage Lemmas 2 and 3 presented after this proof. For part (a), if $r_+(\theta_{m-1}) \leq 0$, it follows from Lemma 2 that $r_+(\widetilde{\theta}_j) \leq 0$ for all $j \in [K-1]$. Lemma 3 implies that the relative difference is non-increasing on $[\theta_{m-1},\theta_m]$ and thus, the maximum relative difference is attained at θ_{m-1}. For part (b), if $r_-(\theta_m) \geq 0$, it follows from Lemma 2 that $r_-(\widetilde{\theta}_j) \geq 0$ for all $j \in [K-1]$. Lemma 3 implies that the relative difference is non-decreasing on $[\theta_{m-1},\theta_m]$ and thus, the maximum relative difference is attained at θ_m. For part (c), if there exists $j' \in [K-1]$ such that j' the first index for which $r_+(\widetilde{\theta}_{j'}) \leq 0$, i.e., $r_+(\widetilde{\theta}_j) > 0$ for all $j < j'$, it follows from Lemma 2 that $r_+(\widetilde{\theta}_j) \leq 0$ for all $j \geq j'$. Lemma 3 implies that the relative difference is non-decreasing on $[\theta_{m-1},\widetilde{\theta}_{j'}]$ and non-increasing on $[\widetilde{\theta}_{j'},\widetilde{\theta}_m]$ and thus, the maximum relative difference is attained at $\widetilde{\theta}_{j'}$. □

Lemma 2. *Assume that $\widehat{v} > 0$. Let $\widetilde{\theta}_0 = \theta_{m-1}$ and $\{\widetilde{\theta}_1,\ldots,\widetilde{\theta}_{K-1}\} \subset (\theta_{m-1},\theta_m)$ be the kink points of the upper bounding function $\widetilde{v}$ on (θ_{m-1},θ_m) with $\widetilde{\theta}_1 < \widetilde{\theta}_2 < \cdots < \widetilde{\theta}_{K-1}$. Then, we have $r_+(\widetilde{\theta}_j) < r_+(\widetilde{\theta}_{j-1})$ for all $j \in [K-1]$.*

Proof. Proof of Lemma 2. For $j \in [K-1]$, we have

$$\begin{aligned}
r_+(\widetilde{\theta}_{j-1}) &= -p_m\widetilde{v}(\widetilde{\theta}_{j-1}) + s_+(\widetilde{\theta}_{j-1})\widehat{v}(\widetilde{\theta}_{j-1}) \\
&= -p_m\big[\widetilde{v}(\widetilde{\theta}_j) - s_+(\widetilde{\theta}_{j-1})(\widetilde{\theta}_j - \widetilde{\theta}_{j-1})\big] + s_+(\widetilde{\theta}_{j-1})\big[\widehat{v}(\widetilde{\theta}_j) - p_m(\widetilde{\theta}_j - \widetilde{\theta}_{j-1})\big] \qquad \text{(A.4)}
\end{aligned}$$

$$
\begin{aligned}
&= \big[-p_m \widetilde{v}(\widetilde{\theta}_j) + s_+(\widetilde{\theta}_{j-1}) \widehat{v}(\widetilde{\theta}_j) \big] + p_m (\widetilde{\theta}_j - \widetilde{\theta}_{j-1}) \big[s_+(\widetilde{\theta}_{j-1}) - s_+(\widetilde{\theta}_{j-1}) \big] \\
&= -p_m \widetilde{v}(\widetilde{\theta}_j) + s_+(\widetilde{\theta}_{j-1}) \widehat{v}(\widetilde{\theta}_j) \\
&> -p_m \widetilde{v}(\widetilde{\theta}_j) + s_+(\widetilde{\theta}_j) \widehat{v}(\widetilde{\theta}_j) = r_+(\widetilde{\theta}_j). \qquad \text{(A.5)}
\end{aligned}
$$

Here, (A.4) follows from the fact that $\widetilde{v}$ is linear on $[\widetilde{\theta}_{j-1}, \widetilde{\theta}_j]$ with slope $s_+(\widetilde{\theta}_{j-1})$ and $\widehat{v}$ is linear on $[\widetilde{\theta}_{j-1}, \widetilde{\theta}_j]$ with slope p_m; (A.5) follows from $\widehat{v}(\widetilde{\theta}_j) > 0$ and the concavity of $\widetilde{v}$ that $s_+(\widetilde{\theta}_j) < s_+(\widetilde{\theta}_{j-1})$. □

Lemma 3. *Consider two lines L_1 and L_2 defined on an interval $[\theta_a, \theta_b] \subseteq [0,1]$ with $L_2 \geq L_1 > 0$. Suppose that the slopes of L_1 and L_2 are p and q, respectively. Let $R(\theta) = [L_2(\theta) - L_1(\theta)]/L_2(\theta)$ be the relative difference between L_1 and L_2 for a given $\theta \in [\theta_a, \theta_b]$. Then, $R(\theta)$ is increasing/remains unchanged/is decreasing on $[\theta_a, \theta_b]$ if and only if the value $-pL_2(\theta_a) + qL_1(\theta_a)$ is greater than 0/equal to 0/less than 0. Moreover, $-pL_2(\theta) + qL_1(\theta) = -pL_2(\theta_a) + qL_1(\theta_a)$ for any $\theta \in [\theta_a, \theta_b]$.*

Proof. Proof of Lemma 3. For $\theta \in (\theta_a, \theta_b)$, we have

$$
\begin{aligned}
\frac{\mathrm{d}R}{\mathrm{d}\theta} &= \frac{\mathrm{d}}{\mathrm{d}\theta} \frac{L_2(\theta) - L_1(\theta)}{L_2(\theta)} \\
&= \frac{\mathrm{d}}{\mathrm{d}\theta} \frac{[L_2(\theta_a) + q(\theta - \theta_a)] - [L_1(\theta_a) + p(\theta - \theta_a)]}{L_2(\theta_a) + q(\theta - \theta_a)} \\
&= \frac{[L_2(\theta_a) + q(\theta - \theta_a)](q - p) - \big\{ [L_2(\theta_a) + q(\theta - \theta_a)] - [L_1(\theta_a) + p(\theta - \theta_a)] \big\} q}{[L_2(\theta_a) + q(\theta - \theta_a)]^2} \\
&= \frac{-pL_2(\theta_a) + qL_1(\theta_a)}{[L_2(\theta_a) + q(\theta - \theta_a)]^2}. \qquad \text{(A.6)}
\end{aligned}
$$

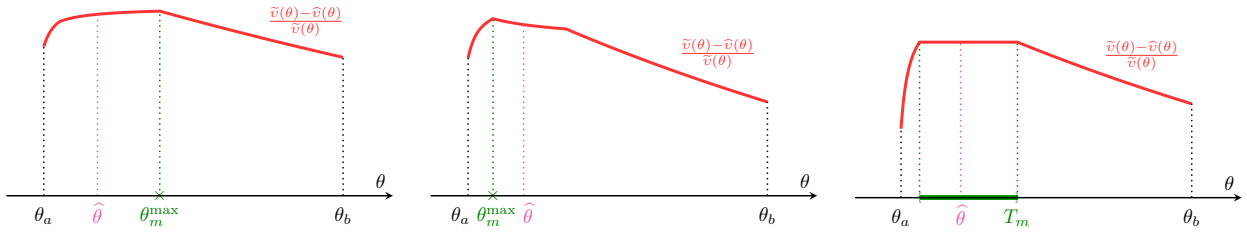
Note that the right derivative of R at θ_a and the left derivative of R at θ_b are the same as (A.6). Since the denominator in (A.6) is always positive, $R(\theta)$ is increasing/remains unchanged/is decreasing on $[\theta_a, \theta_b]$ if and only if the value $-pL_2(\theta_a) + qL_1(\theta_a)$ is greater than 0/equal to 0/less than 0. Finally, for any $\theta \in [\theta_a, \theta_b]$, we have

$$
-pL_2(\theta) + qL_1(\theta) = -p\big[L_2(\theta_a) + q(\theta - \theta_a)\big] + q\big[L_1(\theta_a) + p(\theta - \theta_a)\big] = -pL_2(\theta_a) + qL_1(\theta_a).
$$

This completes the proof. □

Appendix A.6. Proof of Theorem 2

Proof. Proof. To prove the correctness, it suffices to show that $\theta_m^{\max} \in T_m := \arg\max_{\theta \in [\theta_{m-1}, \theta_m]} \big\{ |\widetilde{v}(\theta) - \widehat{v}(\theta)| / \widetilde{v}(\theta) \big\}$ for $m \in \{2, \ldots, M\}$. Consider any $m \in \{2, \ldots, M\}$. Recall that p_m is the slope of the lower bounding function $\widehat{v}$ on $[\theta_{m-1}, \theta_m]$. Also, the upper bounding function $\widetilde{v}$ is piecewise linear and concave on $[\theta_{m-1}, \theta_m]$ (Proposition 3). Therefore, in step 1, if $r_+(\theta_{m-1}) \leq 0$, it follows from part (a) of Proposition 4 that $\theta_m^{\max} = \theta_{m-1} \in T_m$. Also, if $r_-(\theta_m) \geq 0$, it follows from part (b) of Proposition 4 that $\theta_m^{\max} = \theta_m \in T_m$.



(a) $r_+(\widehat{\theta}) > 0$ (b) $r_-(\widehat{\theta}) < 0$ (c) $r_+(\widehat{\theta}) \le 0 \le r_-(\widehat{\theta})$

Figure A.10: An illustration of the three cases in the proof of Theorem 2. The red solid curve represents the relative difference on $[\theta_a, \theta_b]$.

Now, suppose that $r_-(\theta_m) < 0 < r_+(\theta_{m-1})$. In this case, we proceed to the iterative procedure in step 2. We show that in each iteration of step 2, for a given interval (θ_a, θ_b), we will either reach the termination step with $\widehat{\theta} \in T_m$ (computed in step 2a) *or* produce a new interval (θ'_a, θ'_b) such that $T_m \subseteq (\theta'_a, \theta'_b)$, $r_-(\theta'_b) < 0$, and $r_+(\theta'_a) > 0$. First, note that in the first iteration of step 2, we have $\theta_a = \theta_{m-1}$ and $\theta_b = \theta_m$. Since $r_-(\theta_m) < 0 < r_+(\theta_{m-1})$, the initial interval $(\theta_a, \theta_b) = (\theta_{m-1}, \theta_m)$ satisfies $T_m \subseteq (\theta_a, \theta_b)$, $r_-(\theta'_b) < 0$ and $r_+(\theta'_a) > 0$.

Next, assume that the interval (θ_a, θ_b) in the beginning of an iteration of step 2 satisfies $T_m \subseteq (\theta_a, \theta_b)$, $r_-(\theta'_b) < 0$ and $r_+(\theta'_a) > 0$. Recall that $\widehat{\theta}$ in (10) is the value of θ where the two linaer pieces $s_+(\theta_a) \cdot (\theta - \theta_a) + \widetilde{v}(\theta_a)$ and $s_-(\theta_b) \cdot (\theta - \theta_b) + \widetilde{v}(\theta_b)$ of $\widetilde{v}$ intersect. Consider the following three cases; see Figure A.10 for an illustration.

(a) If $r_+(\widehat{\theta}) > 0$, it follows from Lemmas 2 and 3 that the new interval $(\theta'_a, \theta'_b) = (\widehat{\theta}, \theta_b)$ obtained in step 2b satisfies $T_m \subseteq (\theta'_a, \theta'_b)$. Moreover, we have $r_+(\theta'_a) = r_+(\widehat{\theta}) > 0$ and $r_-(\theta'_b) = r_-(\theta_b) < 0$.

(b) If $r_-(\widehat{\theta}) < 0$, it follows from Lemmas 2 and 3 that the new interval $(\theta'_a, \theta'_b) = (\theta_a, \widehat{\theta})$ obtained in step 2b satisfies $T_m \subseteq (\theta'_a, \theta'_b)$. Moreover, we have $r_+(\theta'_a) = r_+(\theta_a) > 0$ and $r_-(\theta'_b) = r_-(\widehat{\theta}) < 0$.

(c) If $r_+(\widehat{\theta}) \le 0 \le r_-(\widehat{\theta})$, it follows from Lemmas 2 and 3 that $\theta_m^{\max} = \widehat{\theta} \in T_m$.

Combining all the three cases, we show that in each iteration within step 2, for a given interval (θ_a, θ_b), we will either reach the termination step with $\widehat{\theta} \in T_m$ or produce a new interval (θ'_a, θ'_b) such that $T_m \subseteq (\theta'_a, \theta'_b)$, $r_-(\theta'_b) < 0$, and $r_+(\theta'_a) > 0$.

Finally, to complete the proof, it suffices to show that the iterative procedure in step 2 terminates in a finite number of iterations. First, we claim that in step 2, for a given interval (θ_a, θ_b), if a new (refined) interval (θ'_a, θ'_b) is obtained in step 2b, then the number of linear pieces of $\widetilde{v}$ on (θ'_a, θ'_b) is strictly less than that of $\widetilde{v}$ on (θ_a, θ_b). To prove this claim, note that for a given interval (θ_a, θ_b), if a refined interval (θ'_a, θ'_b) is obtained in step 2b, then we have $r_+(\widehat{\theta}) < r_+(\theta_a)$ or $r_-(\widehat{\theta}) > r_-(\theta_b)$. Otherwise, we would have $r_+(\widehat{\theta}) \ge r_+(\theta_a) > 0 > r_-(\theta_b) \ge r_-(\widehat{\theta})$, which contradicts with the fact that $r_+(\widehat{\theta}) \le r_-(\widehat{\theta})$ (see Lemma 2). Consider the following two cases; see Figure A.10 for an illustration.

(a) Suppose that $r_+(\widehat{\theta}) < r_+(\theta_a)$. If follows from Lemma 3 that there exists at least one linear

piece of the upper bounding function $\widetilde{v}$ on $[\theta_a, \widehat{\theta}]$ which is not included on $[\widehat{\theta}, \theta_b]$. Thus, the number of linear pieces of $\widetilde{v}$ on the refined interval $(\theta'_a, \theta'_b) = (\widehat{\theta}, \theta_b)$ is strictly less than that of $\widetilde{v}$ on (θ_a, θ_b).

(b) Suppose that $r_-(\widehat{\theta}) > r_-(\theta_b)$. If follows from Lemma 3 that there exists at least one linear piece of the upper bounding function $\widetilde{v}$ on $[\widehat{\theta}, \theta_b]$ which is not included on $[\theta_a, \widehat{\theta}]$. Thus, the number of linear pieces of $\widetilde{v}$ on the refined interval $(\theta'_a, \theta'_b) = (\theta_a, \widehat{\theta})$ is strictly less than that of $\widetilde{v}$ on (θ_a, θ_b).

Combining these two cases, we show that if a new (refined) interval (θ'_a, θ'_b) is obtained in step 2b, then the number of linear pieces of $\widetilde{v}$ on (θ'_a, θ'_b) is strictly less than that of $\widetilde{v}$ on (θ_a, θ_b). Therefore, the number of linear pieces on (θ_a, θ_b) decreases by at least one after each iteration. Also, note that if $\widetilde{v}$ consist of only two linear pieces on (θ_a, θ_b), the iterative procedure must terminate. Consequently, since $\widetilde{v}$ is piecewise linear with a finite number of linear pieces, the iterative procedure in step 2 must terminate in a finite number of iterations. This completes the proof. □

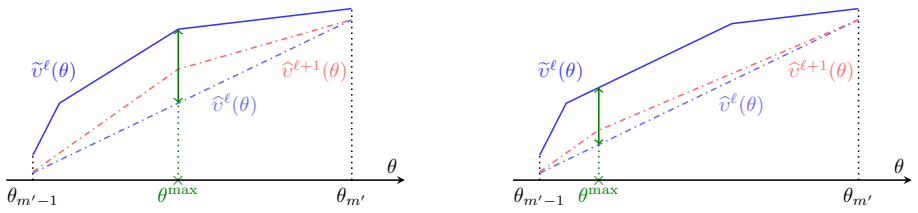
Appendix A.7. Proof of Theorem 3

Proof. Proof. Suppose, on the contrary, that Algorithm 2 does not terminate in a finite number of iterations. Consider any iteration $\ell > 1$, and let $\theta^{\max}$ be the newly identified TRO parameter in iteration $\ell - 1$. In step 1 of iteration ℓ, we solve the TRO-CFLP model with $\theta^{\max}$ and obtain optimal solution $\widehat{\boldsymbol{o}}(\theta^{\max})$. The following are the possible outcomes: (a) $\widehat{\boldsymbol{o}}(\theta^{\max}) \notin \{\widehat{\boldsymbol{o}}(\theta) \mid \theta \in \Theta_{\ell-1}\}$, (b) $\widehat{\boldsymbol{o}}(\theta^{\max}) \in \{\widehat{\boldsymbol{o}}(\theta) \mid \theta \in \Theta_{\ell-1}\}$ and $\acute{\varphi}^{\mathrm{DRO}}(\theta^{\max}) < \acute{\varphi}^{\mathrm{DRO}}(\theta')$ for any $\theta' \in \Theta_{\ell-1}$ such that $\widehat{\boldsymbol{o}}(\theta') = \widehat{\boldsymbol{o}}(\theta^{\max})$, and (c) $\widehat{\boldsymbol{o}}(\theta^{\max}) \in \{\widehat{\boldsymbol{o}}(\theta) \mid \theta \in \Theta_{\ell-1}\}$ and there exists $\theta' \in \Theta_{\ell-1}$ such that $\widehat{\boldsymbol{o}}(\theta') = \widehat{\boldsymbol{o}}(\theta^{\max})$ and $\acute{\varphi}^{\mathrm{DRO}}(\theta^{\max}) = \acute{\varphi}^{\mathrm{DRO}}(\theta')$. Next, we analyze these cases.

Since the set of feasible solutions, $\boldsymbol{o} \in \mathcal{O}$, is finite, case (a) can occur only a finite number of times. For case (b), we obtain a feasible solution $\widehat{\boldsymbol{o}}(\theta^{\max})$ that is the same as $\widehat{\boldsymbol{o}}(\theta)$ for some $\theta \in \Theta_{\ell-1}$. In particular, $\acute{\varphi}^{\mathrm{DRO}}(\theta^{\max}) < \acute{\varphi}^{\mathrm{DRO}}(\theta')$ for any $\theta' \in \Theta_{\ell-1}$ such that $\widehat{\boldsymbol{o}}(\theta') = \widehat{\boldsymbol{o}}(\theta^{\max})$. By step 1d, this occurs only when solution $\widehat{\boldsymbol{o}}(\theta^{\max})$ is obtained for the second time. Since the set of feasible solutions $\boldsymbol{o} \in \mathcal{O}$ is finite, case (b) can occur only a finite number of times. It follows that Algorithm 2 runs indefinitely if case (c) occurs infinitely often.

Let L_A and L_B be the iteration indices such that case (a) and case (b) happen for the last time, respectively. We define $L = \max\{L_A, L_B\}$. From iteration $L + 1$ onward, Algorithm 2 proceeds indefinitely with case (c). That is, from iteration $L + 1$ onward, solutions $\widehat{\boldsymbol{o}}(\theta^{\max})$ (obtained in step 1b) appears for at least three times. It follows from Lemma 4 (presented after this proof) that the upper bounding function is not improved after iteration L (i.e., $\widetilde{v}^{\ell} = \widetilde{v}^{L+1}$ for all $\ell > L$). Next, we show that the lower bounding function keeps improving from iterations $L + 1$ onward until $r^{\max} \leq \varepsilon$ in a finite number of iterations, which contracts the assumption that Algorithm 2 does not terminate in a finite number of iterations.

Sort and enumerate the set Θ^{L+1} as $\{\theta_m\}_{m=1}^{M}$ such that its elements satisfy $0 = \theta_1 < \theta_2 <$



(a) $\theta^{\max}$ at a kink point

(b) $\theta^{\max}$ at a non-kink point

Figure A.11: An illustration of the two cases where the maximum relative difference attained on $[\theta_{m-1}, \theta_m]$ is $r_{m'}^{\max} > \varepsilon$. For case (a), the upper bounding function $\widetilde{v}^{\ell}$ consists of $K_1 = 2$ linear pieces on $[\theta_{m'-1}, \theta^{\max}]$ and $K_2 = 1$ linear piece on $[\theta^{\max}, \theta_{m'}]$. For case (b), the upper bounding function $\widetilde{v}^{\ell}$ consists of $K_1 = K_2 = 2$ linear pieces on both $[\theta_{m'-1}, \theta^{\max}]$ and $[\theta^{\max}, \theta_{m'}]$.

$\cdots < \theta_M = 1$. On each interval $[\theta_{m-1}, \theta_m]$, we have $r_m^{\max} \leq \varepsilon$ or $r_m^{\max} > \varepsilon$ for $m \in \{2, \ldots, M\}$. Let $\mathcal{M} = \{m \in \{2, \ldots, M\} \mid r_m^{\max} > \varepsilon\}$ be the set of indices corresponding to intervals with a maximum relative difference between $\widehat{v}^{L+1}$ and $\widetilde{v}^{L+1}$ greater than ε. Recall that in step 4, we identify a new TRO parameter $\theta^{\max}$ on one of these intervals $\{[\theta_{m-1}, \theta_m]\}_{m \in \mathcal{M}}$. The TRO-CFLP model is then solved with $\theta^{\max}$ in the next iteration. Since we assume that Algorithm 2 does not terminate in a finite number of iterations, it follows that there exists an interval $[\theta_{m'-1}, \theta_{m'}]$ for some $m' \in \mathcal{M}$ that contains an infinite number of new TRO parameters obtained in step 4 from iteration $L+1$ onward.

Now, consider an iteration $\ell > L$ where a new TRO parameter $\theta^{\max}$ obtained in step 4 lies on the interval $[\theta_{m'-1}, \theta_{m'}]$ for the first time after iteration L. There are two possibilities: (i) $\theta^{\max}$ is at a kink point of the upper bounding function $\widetilde{v}^{\ell}$ and (ii) $\theta^{\max}$ is at a non-kink point of $\widetilde{v}^{\ell}$; see Figure A.11 for an illustration. Recall that the upper bounding function is not improved after iteration L, and assume that there are $K \geq 2$ linear pieces on $[\theta_{m'-1}, \theta_{m'}]$. (Note that $K > 1$ since we identify a new TRO parameter $\theta^{\max}$ that lies on $(\theta_{m'-1}, \theta_{m'})$.) Thus, at iteration $\ell + 1$, only the lower bounding function is (strictly) improved. Moreover, after adding $\theta^{\max}$ into the set Θ_ℓ, the interval $[\theta_{m'-1}, \theta_{m'}]$ is partitioned into two parts: $[\theta_{m'-1}, \theta^{\max}]$ and $[\theta^{\max}, \theta_{m'}]$. Suppose that $\widetilde{v}^{\ell}$ consists of K_1 and K_2 linear pieces on intervals $[\theta_{m'-1}, \theta^{\max}]$ and $[\theta^{\max}, \theta_{m'}]$, respectively. Then, we have the following three observations from cases (i) and (ii) above.

(O1) We have $K_1 < K$ and $K_2 < K$ with $K_1 + K_2 \leq K + 1$; see Figure A.11.

(O2) Since $\theta^{\max}$ lies between the first and the last kink points of $\widetilde{v}^{\ell}$ (see Algorithm 1), if either $K_1 = 1$ or $K_2 = 1$, we must have $K_1 + K_2 = K$.

(O3) If $\widetilde{v}^{\ell}$ consists of only one linear piece on a given interval, say $[\theta^{\max}, \theta_{m'}]$, the maximum relative difference between the bounding functions on $[\theta^{\max}, \theta_{m'}]$ is attained at either $\theta^{\max}$ or $\theta_{m'}$. Since the relative differences at both $\theta^{\max}$ and $\theta_{m'}$ are less than or equal to ε, the new TRO parameters obtained in step 4 in subsequent iterations do not belong to this interval.

With these observations, we claim that there are at most $2K - 3$ new TRO parameters on

$[\theta_{m'-1}, \theta_{m'}]$ obtained at step 4, where $K < \infty$. We prove this claim by induction on K. It is trivial that when $K = 2$, i.e., when $\widetilde{v}^\ell$ consists of only two linear pieces on $[\theta_{m'-1}, \theta_{m'}]$, only one new TRO parameter will be obtained in step 4. Assume that the claim holds for any $K \leq k-1$, where $k > 2$. Consider the case when $K = k$ (induction step), i.e., when $\widetilde{v}^\ell$ consists of k linear pieces on $[\theta_{m'-1}, \theta_{m'}]$. After obtaining the first TRO parameter $\theta^{\max}$ on $[\theta_{m'-1}, \theta_{m'}]$ in step 4, $\widetilde{v}^\ell$ consists of k_1 and k_2 linear pieces on the two intervals $[\theta_{m'-1}, \theta^{\max}]$ and $[\theta^{\max}, \theta_{m'}]$, respectively. It follows from observation (O1) that $k_1 < k$ and $k_2 < k$ with $k_1 + k_2 \leq k + 1$.

- Suppose that $k_1 \geq 2$ and $k_2 \geq 2$. Then, by the induction assumption, there is at most $2k_1 - 3$ and $2k_2 - 3$ new TRO parameters obtained at step 4 on these two intervals. Thus, the number of new TRO parameters obtained in step 4 after iteration L (including ℓ) is upper bounded by $1 + (2k_1 - 3) + (2k_2 - 3) = 2(k_1 + k_2) - 5 \leq 2(k + 1) - 5 = 2k - 3$.
- Suppose that either $k_1 = 1$ or $k_2 = 1$. Without loss of generality, let $k_1 = 1$. If follows from observation (O2) that $k_2 = k - 1$. Since $k_1 = 1$, by observation (O3), no new TRO parameters on $[\theta_{m'-1}, \theta^{\max}]$ are obtained. By the induction assumption, there is at most $2(k-1)-3 = 2k-5$ new TRO parameters on $[\theta^{\max}, \theta_{m'}]$ that will be obtained in step 4. Thus, the number of new TRO parameters that will be obtained in step 4 after iteration L (including ℓ) is upper bounded by $1 + (2k - 5) = 2k - 4 \leq 2k - 3$.

This completes the induction step. Therefore, there is only a finite number of new TRO parameters that will be obtained in step 4 on any interval $[\theta_{m'-1}, \theta_{m'}]$ for $m' \in \mathcal{M}$, which gives the desired contradiction. Thus, Algorithm 2 terminates in a finite number of iterations. □

Lemma 4. *Consider iteration $h_1 > 1$ of Algorithm 2. Suppose that the best feasible solution $\overline{\boldsymbol{o}} \in \mathcal{O}$ obtained by solving the TRO-CFLP model in step 1b has not been obtained in previous iterations. Consider the first subsequent iteration $h_2 > h_1$, where the best feasible solution obtained by solving the TRO-CFLP model in step 1b is the same as $\overline{\boldsymbol{o}}$. Consider the TRO parameter θ^{max} identified in iteration $k \geq h_2$. If the best feasible solution obtained from solving the TRO-CFLP model with θ^{max} in step 1b of iteration $k+1$ is $\overline{\boldsymbol{o}}$, we have $\widetilde{v}^{k+1} = \widetilde{v}^k$, i.e., the upper bounding function in iteration $k+1$ is the same as the upper bounding function in iteration k.*

Proof. Proof of Lemma 4.

Let $\theta^{\max}_{\ell+1}$ be the value of $\theta^{\max}$ obtained in step 4 of iteration $\ell \in \mathbb{N}$ with $\theta^{\max}_0 = 0$ and $\theta^{\max}_1 = 1$. Let $\widehat{\varphi}^{\mathrm{SP}}_\ell := \widehat{\varphi}^{\mathrm{SP}}(\theta^{\max}_\ell)$ and $\widehat{\varphi}^{\mathrm{DRO}}_\ell := \acute{\varphi}^{\mathrm{DRO}}(\theta^{\max}_\ell)$ for all $\ell \in \mathbb{Z}_+$. Let $\widehat{\boldsymbol{o}}_\ell$ be the best feasible obtained from solving the TRO-CFLP model with $\theta = \theta^{\max}_\ell$ for all $\ell \in \mathbb{Z}_+$. In step 3b of iteration $k+1 > h_2$, we update the upper bounding function using (9) as follows

$$\widetilde{v}^{k+1}(\theta) = \min_{\ell \in \{0,1,\ldots,h_2,\ldots,k+1\}} \left\{ \sum_{i \in I} c^{\mathrm{f}}_i(\widehat{\boldsymbol{o}}_\ell)_i + (1-\theta)\widehat{\varphi}^{\mathrm{SP}}_\ell + \theta \acute{\varphi}^{\mathrm{DRO}}_\ell \right\} =: \min_{\ell \in \{0,1,\ldots,k+1\}} \left\{ L_\ell(\theta) \right\}.$$

By our assumption, in iteration h_1, we identify the solution $\widehat{\boldsymbol{o}}_{h_1} = \overline{\boldsymbol{o}}$ for the first time and in iteration h_2, we again obtain $\widehat{\boldsymbol{o}}_{h_2} = \overline{\boldsymbol{o}}$. Hence, in step 1d of iteration h_2, we set $\acute{\varphi}^{\mathrm{DRO}}_{h_2} = \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\overline{\boldsymbol{o}}, \boldsymbol{d})]$.

Accordingly,

$$\begin{aligned} L_{h_2}(\theta) &= \sum_{i\in I} c^{\mathrm{f}}_i(\widehat{\boldsymbol{o}}_{h_2})_i + (1-\theta)\widehat{\varphi}^{\mathrm{SP}}_{h_2} + \theta\acute{\varphi}^{\mathrm{DRO}}_{h_2} \\ &= \sum_{i\in I} c^{\mathrm{f}}_i(\overline{\boldsymbol{o}})_i + (1-\theta)\mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\overline{\boldsymbol{o}},\boldsymbol{d})] + \theta \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\overline{\boldsymbol{o}},\boldsymbol{d})]. \end{aligned}$$

Since the best feasible solution obtained by solving the TRO-CFLP model in step 1b of iteration $k+1$ is again $\overline{\boldsymbol{o}}$, we have $\widehat{\boldsymbol{o}}_{k+1} = \overline{\boldsymbol{o}}$ and $\acute{\varphi}^{\mathrm{DRO}}_{k+1} = \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\overline{\boldsymbol{o}},\boldsymbol{d})]$. Therefore,

$$\begin{aligned} L_{k+1}(\theta) &= \sum_{i\in I} c^{\mathrm{f}}_i(\widehat{\boldsymbol{o}}_{k+1})_i + (1-\theta)\widehat{\varphi}^{\mathrm{SP}}_{k+1} + \theta\acute{\varphi}^{\mathrm{DRO}}_{k+1} \\ &= \sum_{i\in I} c^{\mathrm{f}}_i(\overline{\boldsymbol{o}})_i + (1-\theta)\mathbb{E}_{\widehat{\mathbb{P}}_N}[Q(\overline{\boldsymbol{o}},\boldsymbol{d})] + \theta \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\overline{\boldsymbol{o}},\boldsymbol{d})]. \end{aligned}$$

This shows that $L_{k+1} \equiv L_{h_2}$ (with $k+1 > h_2$). Therefore,

$$\widetilde{v}^{k+1}(\theta) = \min_{\ell\in\{0,1,\ldots,h_2,\ldots,k+1\}} \big\{L_\ell(\theta)\big\} = \min_{\ell\in\{0,1,\ldots,h_2,\ldots,k\}} \big\{L_\ell(\theta)\big\} = \widetilde{v}^{k}(\theta).$$

This completes the proof. □

Appendix A.8. Proof of Proposition 5

Proof. Proof.

First, consider the primal C&CG optimality cut (12d). Note that $\max_{e\in E}\big\{\sum_{i\in I} C_i v^e_i o_i + \sum_{j\in J} w^e_j d_j\big\} = Q(\boldsymbol{o},\boldsymbol{d})$. Using the definition of $Q(\boldsymbol{o},\boldsymbol{d})$ in (1), we can rewrite (12d) as

$$\delta \geq \min_{(\boldsymbol{x},\boldsymbol{u})\in\mathcal{X}} \left\{ \sum_{i\in I}\sum_{j\in J} t_{i,j}x_{i,j} + \sum_{j\in J} c^{\mathrm{u}}_j u_j \right\} - \sum_{k\in[K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'} g_{k'}(\boldsymbol{d}), \tag{A.7}$$

where $\mathcal{X} = \{(\boldsymbol{x},\boldsymbol{u}) \geq \boldsymbol{0} \mid (1b)\text{–}(1c)\}$. Hence, (A.7) is equivalent to

$$\delta \geq \sum_{i\in I}\sum_{j\in J} t_{i,j}x_{i,j}(\boldsymbol{d}) + \sum_{j\in J} c^{\mathrm{u}}_j u_j(\boldsymbol{d}) - \sum_{k\in[K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'} g_{k'}(\boldsymbol{d}), \tag{A.8a}$$

$$\sum_{j\in J} x_{i,j}(\boldsymbol{d}) \leq C_i o_i, \quad \sum_{i\in I} x_{i,j}(\boldsymbol{d}) + u_j(\boldsymbol{d}) \geq d_j, \quad \forall i\in I,\ j\in J, \tag{A.8b}$$

$$x_{i,j}(\boldsymbol{d}) \geq 0, \quad u_j(\boldsymbol{d}) \geq 0, \quad \forall i\in I,\ j\in J. \tag{A.8c}$$

The equivalence follows from the fact that (A.7) holds if and only if there exists $(\boldsymbol{x},\boldsymbol{u}) \in \mathcal{X}$ such that (A.8a) holds.

Next, consider the dual C&CG optimality cut (12e). We first dualize the convex functions $\{f_k\}_{k\in[K]}$ and $\{g_{k'}\}_{k'\in[K']}$ to reformulate the maximization problem over $\boldsymbol{d}\in\mathcal{S}$ (i.e., $\sup_{\boldsymbol{d}\in\mathcal{S}}\{\cdot\}$) into a minimization problem:

$$\sup_{\boldsymbol{d}\in\mathcal{S}} \left\{ \sum_{j\in J} w^e_j d_j - \sum_{k\in[K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'} g_{k'}(\boldsymbol{d}) \right\}$$

$$= \sup_{\boldsymbol{d}\in\mathcal{S}} \inf_{\boldsymbol{\kappa},\boldsymbol{\tau}} \left\{ \sum_{j\in J} w_j^e d_j - \sum_{k\in[K]} \rho_k \left[\boldsymbol{d}^\top \boldsymbol{\kappa}_k - f_k^*(\boldsymbol{\kappa}_k) \right] - \sum_{k'\in[K']} \gamma_{k'} \left[\boldsymbol{d}^\top \boldsymbol{\tau}_{k'} - g_{k'}^*(\boldsymbol{\tau}_{k'}) \right] \right\} \tag{A.9a}$$

$$= \inf_{\boldsymbol{\kappa},\boldsymbol{\tau}} \left\{ \sum_{k\in[K]} \rho_k f_k^*(\boldsymbol{\kappa}_k) + \sum_{k'\in[K']} \gamma_{k'} g_{k'}^*(\boldsymbol{\tau}_{k'}) + \sup_{\boldsymbol{d}\in\mathcal{S}} \left\{ \sum_{j\in J} w_j^e d_j - \sum_{k\in[K]} \rho_k (\boldsymbol{\kappa}_k^\top \boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'} (\boldsymbol{\tau}_{k'}^\top \boldsymbol{d}) \right\} \right\} \tag{A.9b}$$

$$= \inf_{\boldsymbol{\kappa},\boldsymbol{\tau}} \left\{ \sum_{k\in[K]} \rho_k f_k^*(\boldsymbol{\kappa}_k) + \sum_{k'\in[K']} \gamma_{k'} g_{k'}^*(\boldsymbol{\tau}_{k'}) + \min_{\boldsymbol{\pi}\in\mathcal{D}(\boldsymbol{\kappa},\boldsymbol{\tau})} \left\{ \sum_{j\in J} (\overline{\pi}_j \overline{d}_j - \underline{\pi}_j \underline{d}_j) \right\} \right\}, \tag{A.9c}$$

where $\mathcal{D}(\boldsymbol{\kappa},\boldsymbol{\tau}) = \left\{ (\overline{\boldsymbol{\pi}},\underline{\boldsymbol{\pi}}) \in \mathbb{R}^{|J|} \times \mathbb{R}^{|J|} \mid \boldsymbol{w} - \sum_{k\in[K]} \rho_k \boldsymbol{\kappa}_k - \sum_{k'\in[K']} \gamma_{k'} \boldsymbol{\tau}_{k'} = \overline{\boldsymbol{\pi}} - \underline{\boldsymbol{\pi}} \right\}$. Here, (A.9a) follows from the conjugacy theorem for proper, closed, convex functions $\{f_k\}_{k\in[K]}$ and $\{g_{k'}\}_{k'\in[K']}$ (see Proposition 1.6.1 of Bertsekas, 2009), (A.9b) follows from the Sion min-max theorem (Sion, 1958), and (A.9c) follows from the LP duality of the inner maximization problem in (A.9b). Replacing problem $\sup_{\boldsymbol{d}\in\mathcal{S}}\{\cdot\}$ with its equivalent reformulation (A.9c), we reformulate optimality cut (12e) as

$$\delta \geq \sum_{i\in I} C_i v_i^e o_i + \sum_{k\in[K]} \rho_k f_k^*(\boldsymbol{\kappa}_k^e) + \sum_{k'\in[K']} \gamma_{k'} g_{k'}^*(\boldsymbol{\tau}_{k'}^e) + \sum_{j\in J} (\overline{\pi}_j \overline{d}_j - \underline{\pi}_j \underline{d}_j), \tag{A.10a}$$

$$\boldsymbol{w} - \sum_{k\in[K]} \rho_k \boldsymbol{\kappa}_k^e - \sum_{k'\in[K']} \gamma_{k'} \boldsymbol{\tau}_{k'}^e = \overline{\boldsymbol{\pi}} - \underline{\boldsymbol{\pi}}, \tag{A.10b}$$

$$\boldsymbol{\kappa}_k^e \in \mathbb{R}^{|J|}, \quad \boldsymbol{\tau}_{k'}^e \in \mathbb{R}^{|J|}, \quad \forall k \in [K],\, k' \in [K']. \tag{A.10c}$$

The equivalence follows from the fact (12e) holds if and only if there exist $(\boldsymbol{\kappa}^e,\boldsymbol{\tau}^e) \in \mathbb{R}^{K\times|J|} \times \mathbb{R}^{K'\times|J|}$ and $(\overline{\boldsymbol{\pi}},\underline{\boldsymbol{\pi}}) \in \mathcal{D}(\boldsymbol{\kappa}^e,\boldsymbol{\tau}^e)$ such that (A.10a) holds. □

Appendix A.9. Proof of Proposition 6

Proof. Proof. Let $(\overline{\boldsymbol{d}},\overline{\boldsymbol{w}},\overline{\boldsymbol{v}}) \in \mathcal{S} \times \Pi$ be an optimal solution to the subproblem (13). We first show that $(\overline{\boldsymbol{w}},\overline{\boldsymbol{v}}) \in \arg\max_{(\boldsymbol{w},\boldsymbol{v})\in\Pi} \left\{ \sum_{i\in I} C_i v_i o_i^h + \sum_{j\in J} w_j \overline{d}_j \right\}$. Suppose, on the contrary, that $(\overline{\boldsymbol{w}},\overline{\boldsymbol{v}}) \notin \arg\max_{(\boldsymbol{w},\boldsymbol{v})\in\Pi} \left\{ \sum_{i\in I} C_i v_i o_i^h + \sum_{j\in J} w_j \overline{d}_j \right\}$. Then, there exists $(\boldsymbol{w}^\star,\boldsymbol{v}^\star) \in \Pi$ such that $\sum_{i\in I} C_i v_i^\star o_i^h + \sum_{j\in J} w_j^\star \overline{d}_j > \sum_{i\in I} C_i \overline{v}_i o_i^h + \sum_{j\in J} \overline{w}_j \overline{d}_j$. Therefore, we have

$$\begin{aligned} &\sum_{i\in I} C_i v_i^\star o_i^h + \sum_{j\in J} w_j^\star \overline{d}_j - \sum_{k\in[K]} \rho_k^h f_k(\overline{\boldsymbol{d}}) - \sum_{k'\in[K']} \gamma_{k'}^h g_{k'}(\overline{\boldsymbol{d}}) \\ &> \sum_{i\in I} C_i \overline{v}_i o_i^h + \sum_{j\in J} \overline{w}_j \overline{d}_j - \sum_{k\in[K]} \rho_k^h f_k(\overline{\boldsymbol{d}}) - \sum_{k'\in[K']} \gamma_{k'}^h g_{k'}(\overline{\boldsymbol{d}}), \end{aligned}$$

which contradicts the optimality of $(\overline{\boldsymbol{d}},\overline{\boldsymbol{w}},\overline{\boldsymbol{v}})$. This shows that $(\overline{\boldsymbol{w}},\overline{\boldsymbol{v}}) \in \arg\max_{(\boldsymbol{w},\boldsymbol{v})\in\Pi} \left\{ \sum_{i\in I} C_i v_i o_i^h + \sum_{j\in J} w_j \overline{d}_j \right\}$. Finally, since the dual of the recourse $Q(\boldsymbol{o}^h,\overline{\boldsymbol{d}})$ is an LP, there always exists an optimal extreme point solution $(\boldsymbol{w}^e,\boldsymbol{v}^e) \in \arg\max_{(\boldsymbol{w},\boldsymbol{v})\in\Pi} \left\{ \sum_{i\in I} C_i v_i o_i^h + \sum_{j\in J} w_j \overline{d}_j \right\}$ such that $(\overline{\boldsymbol{d}},\boldsymbol{w}^e,\boldsymbol{v}^e) \in \mathcal{S} \times \Pi$ is an optimal solution to the subproblem (13). □

Appendix A.10. Proof of Proposition 7

Proof. Proof. Note that subproblem (13) is equivalent to

$$\sup_{\boldsymbol{d}\in\mathcal{S}}\left\{Q(\boldsymbol{o},\boldsymbol{d})-\sum_{k\in[K]}\rho_k^h f_k(\boldsymbol{d})-\sum_{k'\in[K']}\gamma_{k'}^h g_{k'}(\boldsymbol{d})\right\} \tag{A.11}$$

To reformulate (A.11), we first derive an equivalent reformulation of $Q(\boldsymbol{o},\boldsymbol{d})$. Note that constraints (1b) in $Q(\boldsymbol{o},\boldsymbol{d})$ always achieve equality at optimality (since we are minimizing the traveling cost and unmet demand penalty). Thus, we can rewrite $Q(\boldsymbol{o},\boldsymbol{d})$ as

$$\underset{\boldsymbol{x}\geq\boldsymbol{0},\boldsymbol{u}\geq\boldsymbol{0}}{\text{minimize}} \quad \sum_{i\in I}\sum_{j\in J}t_{i,j}x_{i,j}+\sum_{j\in J}c_j^{\text{u}}u_j \tag{A.12a}$$

$$\text{subject to} \quad \sum_{j\in J}x_{i,j}\leq C_i o_i, \quad \forall i\in I, \tag{A.12b}$$

$$\sum_{i\in I}x_{i,j}+u_j=d_j, \quad \forall j\in J. \tag{A.12c}$$

The dual of (A.12) is as follows.

$$\underset{\boldsymbol{w},\boldsymbol{v}}{\text{maximize}} \quad \sum_{i\in I}C_i v_i o_i+\sum_{j\in J}w_j d_j \tag{A.13a}$$

$$\text{subject to} \quad w_j+v_i\leq t_{i,j}, \quad \forall i\in I,\, j\in J, \tag{A.13b}$$

$$w_j\leq c_j^{\text{u}}, \quad \forall j\in J, \tag{A.13c}$$

$$v_i\leq 0, \quad \forall i\in I. \tag{A.13d}$$

Note that at optimality, $w_j\geq 0$ for all $j\in J$ since $Q(\boldsymbol{o},\boldsymbol{d})$ is non-decreasing in d_j. Let $(\boldsymbol{x},\boldsymbol{u})$ and $(\boldsymbol{w},\boldsymbol{v})$ be a feasible solution to the primal problem (A.12) and the dual problem (A.13), respectively. The pair $(\boldsymbol{x},\boldsymbol{u})$ is primal optimal and $(\boldsymbol{w},\boldsymbol{v})$ is dual optimal if and only if they satisfy the following complementary slackness conditions:

$$\left(C_i o_i-\sum_{j\in J}x_{i,j}\right)(-v_i)=0, \quad \forall i\in I, \tag{A.14a}$$

$$(t_{i,j}-w_j-v_i)x_{i,j}=0, \quad \forall i\in I,\, j\in J, \tag{A.14b}$$

$$(c_j^{\text{u}}-w_j)u_j=0, \quad \forall j\in J. \tag{A.14c}$$

The set $\{(\boldsymbol{u},\boldsymbol{x},\boldsymbol{w},\boldsymbol{v}) \mid$ (A.12b)–(A.12c), (A.13b)–(A.13c), (A.14a)–(A.14c)$\}$, which consists of the constraints on primal feasibility, dual feasibility, and complimentary slackness conditions, characterize the set of primal and dual optimal solutions. Therefore, we can reformulate the second-stage problem $Q(\boldsymbol{o},\boldsymbol{d})$ as

$$\underset{\boldsymbol{u},\boldsymbol{x},\boldsymbol{w},\boldsymbol{v}}{\text{maximize}} \quad \sum_{i\in I}\sum_{j\in J}t_{i,j}x_{i,j}+\sum_{j\in J}c_j^{\text{u}}u_j \tag{A.15a}$$

$$\text{subject to} \quad \text{(A.12b)–(A.12c)}, \text{(A.13b)–(A.13c)}, \text{(A.14a)–(A.14c)}, \tag{A.15b}$$

$$x_{i,j} \geq 0,\ u_j \geq 0,\ v_i \leq 0, \quad \forall i \in I,\ j \in J. \tag{A.15c}$$

Note that (A.14a)–(A.14c) are non-linear. To linearize, we can introduce a binary variable for each constraint and the following constraints:

$$C_i o_i - \sum_{j\in J} x_{i,j} \leq M_i^{1,1} s_i^1,\ -v_i \leq M_i^{1,2}(1 - s_i^1), \quad \forall i \in I, \tag{A.16a}$$

$$t_{i,j} - w_j - v_i \leq M_{i,j}^{2,1} s_{i,j}^2,\ x_{i,j} \leq M_{i,j}^{2,2}(1 - s_{i,j}^2), \quad \forall i \in I,\ j \in J, \tag{A.16b}$$

$$c_j^{\mathrm{u}} - w_j \leq M_j^{3,1} s_j^3,\ u_j \leq M_j^{3,2}(1 - s_j^3), \quad \forall j \in J, \tag{A.16c}$$

$$s_j^1 \in \{0,1\},\ s_{i,j}^2 \in \{0,1\},\ s_j^3 \in \{0,1\}, \quad \forall i \in I,\ j \in J. \tag{A.16d}$$

Thus, we can reformulate the second-stage problem $Q(\boldsymbol{o}, \boldsymbol{d})$ as

$$\underset{\boldsymbol{u},\boldsymbol{x},\boldsymbol{w},\boldsymbol{v},\boldsymbol{s}}{\text{maximize}} \quad \sum_{i\in I}\sum_{j\in J} t_{i,j} x_{i,j} + \sum_{j\in J} c_j^{\mathrm{u}} u_j \tag{A.17a}$$

$$\text{subject to} \quad \text{(A.12b)–(A.12c)}, \text{(A.13b)–(A.13c)}, \text{(A.16a)–(A.16c)}, \tag{A.17b}$$

$$x_{i,j} \geq 0,\ u_j \geq 0,\ v_i \leq 0,\ s_j^1 \in \{0,1\},\ s_{i,j}^2 \in \{0,1\},\ s_j^3 \in \{0,1\}, \quad \forall i \in I,\ j \in J. \tag{A.17c}$$

Therefore, using the maximization reformulation of $Q(\boldsymbol{o}, \boldsymbol{d})$ in (A.17), we can reformulate the sub-problem (13), which is equivalent to $\sup_{\boldsymbol{d}\in\mathcal{S}} \left\{Q(\boldsymbol{o}, \boldsymbol{d}) - \sum_{k\in[K]} \rho_k^h f_k(\boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'}^h g_{k'}(\boldsymbol{d})\right\}$, as the desired MINLP (16).

To complete the proof, we next derive tight values for the big-M parameters. (i) For $M_i^{1,1}$, note that $C_i o_i \leq C_i$ and $\sum_{j\in J} x_{i,j} \geq 0$. Thus, we can set $M_i^{1,1} = C_i$. (ii) For $M_i^{1,2}$, it follows from Lemma 5 (presented below) that $-v_i \leq \max_{j\in J}\{c_j^{\mathrm{u}}\} - \min_{j\in J}\{t_{i,j}\}$. Thus, we can set $M_i^{1,2} = \max_{j\in J}\{c_j^{\mathrm{u}}\} - \min_{j\in J}\{t_{i,j}\}$. (iii) For $M_{i,j}^{2,1}$, note that $w_j \geq 0$ and $v_i \geq -\max_{j\in J}\{c_j^{\mathrm{u}}\} + \min_{j\in J}\{t_{i,j}\}$. Thus, we can set $M_{i,j}^{2,1} = t_{i,j} + M_i^{1,2}$. (iv) For $M_{i,j}^{2,2}$, since $x_{i,j} \leq C_i$ by (A.12b), we can set $M_{i,j}^{2,2} = C_i$. (v) For $M_j^{3,1}$, since $w_j \geq 0$, we can set $M_j^{3,1} = c_j^{\mathrm{u}}$. (vi) For $M_j^{3,2}$, note that $u_j \leq d_j \leq \overline{d}_j$ by (A.12c). Thus, we can set $M_j^{3,2} = \overline{d}_j$. This completes the proof. □

Lemma 5. *The following bound is a valid lower bound on the dual variables v_i in (A.13):*

$$v_i \geq -\max_{j\in J}\{c_j^{\mathrm{u}}\} + \min_{j\in J}\{t_{i,j}\}, \quad \forall i \in I. \tag{A.18}$$

Proof. Proof of Lemma 5. For notational simplicity, we use $a \wedge b$ to denote $\min\{a, b\}$ for any $\{a, b\} \subset \mathbb{R}$. We first claim that for any extreme point $(\overline{\boldsymbol{w}}, \overline{\boldsymbol{v}})$ of the dual feasible set $\Pi = \{(\boldsymbol{w}, \boldsymbol{v}) \mid w_j + v_i \leq t_{i,j},\ 0 \leq w_j \leq c_j^{\mathrm{u}},\ v_i \leq 0,\ \forall i \in I,\ j \in J\}$, either $\overline{v}_i = 0$ or $\overline{v}_i = t_{i,j} - \overline{w}_j$ for some $j \in J$. Suppose, on the contrary, that there exists $i \in I$ such that $\overline{v}_i < \min_{j\in J}\left\{t_{i,j} - \overline{w}_j\right\} \wedge 0$. Consider $(\overline{\boldsymbol{w}}, \tilde{\boldsymbol{v}}^1)$ and $(\overline{\boldsymbol{w}}, \tilde{\boldsymbol{v}}^2)$, where

$$\tilde{v}_{i'}^1 = \begin{cases} \overline{v}_{i'} & \text{if } i' \neq i, \\ v' & \text{if } i' = i \end{cases} \quad \text{and} \quad \tilde{v}_{i'}^2 = \begin{cases} \overline{v}_{i'} & \text{if } i' \neq i, \\ \min_{j\in J}\left\{t_{i,j} - \overline{w}_j\right\} \wedge 0 & \text{if } i' = i \end{cases}$$

for some $v' < \overline{v}_i$. Note that, by construction, both $(\overline{\boldsymbol{w}}, \tilde{\boldsymbol{v}}^1)$ and $(\overline{\boldsymbol{w}}, \tilde{\boldsymbol{v}}^2)$ are feasible. Indeed, the dual constraints $w_j + v_i \leq t_{i,j}$ on $(\boldsymbol{w}, \boldsymbol{v})$ trivially holds for all $j \in J$ if $i' \neq i$. Moreover, when $i' = i$, we have

$$\overline{w}_j + \tilde{v}_i^1 = \overline{w}_j + v' < \overline{w}_j + \overline{v}_i \leq t_{i,j}$$

and

$$\overline{w}_j + \tilde{v}_i^2 = \overline{w}_j + \left(\min_{j' \in J} \left\{ t_{i,j'} - \overline{w}_{j'} \right\} \wedge 0 \right) \leq \overline{w}_j + \left(t_{i,j} - \overline{w}_j \right) = t_{i,j}$$

for all $j \in J$. Since $(\overline{\boldsymbol{w}}, \overline{\boldsymbol{v}}) = (1-\lambda)(\overline{\boldsymbol{w}}, \tilde{\boldsymbol{v}}^1) + \lambda(\overline{\boldsymbol{w}}, \tilde{\boldsymbol{v}}^2)$ with $\lambda = (\overline{v}_i - v') / \left[\left(\min_{j \in J} \left\{ t_{i,j} - \overline{w}_j \right\} \wedge 0 \right) - v' \right] \in (0,1)$, this contradicts that $(\overline{\boldsymbol{w}}, \overline{\boldsymbol{v}})$ is an extreme point of Π. Thus, we must have either $\overline{v}_i = 0$ or $\overline{v}_i = t_{i,j} - \overline{w}_j$ for some $j \in J$.

Next, we claim that (A.18) is not violated by any extreme points of Π. Let us consider the following two cases. First, when $\overline{v}_i = 0$, we have

$$- \max_{j' \in J} \{ c_{j'}^{\mathrm{u}} \} + \min_{j' \in J} \{ t_{i,j'} \} < 0 = \overline{v}_i,$$

where the strict inequality follows from $c_j^{\mathrm{u}} > t_{i,j}$ for all $i \in I$ and $j \in J$. Second, when $\overline{v}_i = t_{i,j} - \overline{w}_j$ for some $j \in J$, we have

$$- \max_{j' \in J} \{ c_{j'}^{\mathrm{u}} \} + \min_{j' \in J} \{ t_{i,j'} \} \leq -c_j^{\mathrm{u}} + t_{i,j} \leq -\overline{w}_j + t_{i,j} = \overline{v}_i,$$

where the last inequality follows from (A.13c) since $(\overline{\boldsymbol{w}}, \overline{\boldsymbol{v}}) \in \Pi$. Thus, any extreme point $(\overline{\boldsymbol{w}}, \overline{\boldsymbol{v}})$ of Π satisfies (A.18). This completes the proof. □

Appendix A.11. Proof of Lemma 1

Proof. Proof. Note that $\sup_{\boldsymbol{o} \in \mathcal{O}} \sup_{\boldsymbol{d} \in \mathcal{S}} |Q(\boldsymbol{o}, \boldsymbol{d})| < \infty$ since $\mathcal{O}$ is finite, $\mathcal{S}$ is compact, and $|Q(\boldsymbol{o}, \boldsymbol{d})| < \infty$ for all $\boldsymbol{d} \in \mathcal{S}$. Moreover, by Remark 2.1 in Zhang et al. (2016), the Slater-type condition (18) implies that

$$\left\{ \boldsymbol{\rho} \in \mathbb{R}^K, \boldsymbol{\gamma} \in \mathbb{R}_+^K \;\middle|\; - \sum_{k \in [K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k' \in [K']} \gamma_{k'} g_{k'}(\boldsymbol{d}) \leq 0, \quad \forall \boldsymbol{d} \in \mathcal{S} \right\} = \{ \boldsymbol{0} \}.$$

Therefore, it follows from Proposition 2.4 of Xu et al. (2018) that the set of optimal dual solutions to (A.3) is uniformly bounded. □

Appendix A.12. Proof of Theorem 4

Proof. Proof.

To facilitate the subsequent technical discussions, we use the following generic form of the optimality cuts in the master problem $\delta \geq q(\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}; \boldsymbol{c})$ for all $\boldsymbol{c} \in \overline{\mathcal{C}} \subseteq \mathcal{C}$ with

$$\sup_{\boldsymbol{c} \in \mathcal{C}} q(\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}; \boldsymbol{c}) = \sup_{\boldsymbol{d} \in \mathcal{S},\, (\boldsymbol{w}, \boldsymbol{v}) \in \Pi} \left\{ \sum_{i \in I} C_i o_i v_i + \sum_{j \in J} w_j d_j - \sum_{k \in [K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k' \in [K']} \gamma_{k'} g_{k'}(\boldsymbol{d}) \right\}.$$

Specifically, for BD, we have $\mathcal{C} = \mathcal{S} \times E$ with $q(\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}; \boldsymbol{c})$ defined in (17); for pC&CG, we have $\mathcal{C} = \mathcal{S}$ with $q(\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}; \boldsymbol{c})$ defined in (12d); for dC&CG, we have $\mathcal{C} = E$ with $q(\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}; \boldsymbol{c})$ defined in (12e); for hC&CG, we have $\mathcal{C} = \mathcal{S} \times E$ with $q(\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}; \boldsymbol{c})$ defined as the maximum of the two optimality cuts in (12d) and (12e). In Lemma 6, we prove the finite convergence of the decomposition algorithm using this generic form of optimality cut, i.e., Algorithm 3 with (12d) and (12e) replaced by

$$\delta \geq q(\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}; \boldsymbol{c}), \quad \forall \boldsymbol{c} \in \overline{\mathcal{C}}. \tag{A.19}$$

Then, we will apply Lemma 6 to show the finite convergence of the BD, pC&CG, dC&CG, and hC&CG algorithms.

Lemma 6. *Let $\boldsymbol{y} = (\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \delta)$. Assume that the function $q'(\boldsymbol{y}; \boldsymbol{c}) := q(\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}; \boldsymbol{c}) - \delta$ is uniformly Lipschitz continuous in $\boldsymbol{y}$ over $\boldsymbol{c} \in \mathcal{C}$, i.e., there exists $L > 0$ such that for any $\boldsymbol{y}_1 = (\boldsymbol{o}_1, \boldsymbol{\rho}_1, \boldsymbol{\gamma}_1, \delta_1)$ and $\boldsymbol{y}_2 = (\boldsymbol{o}_2, \boldsymbol{\rho}_2, \boldsymbol{\gamma}_2, \delta_2)$, we have*

$$\left| q'(\boldsymbol{y}_1; \boldsymbol{c}) - q'(\boldsymbol{y}_2; \boldsymbol{c}) \right| \leq L \|\boldsymbol{y}_1 - \boldsymbol{y}_2\|, \quad \forall \boldsymbol{c} \in \mathcal{C}. \tag{A.20}$$

If $\varepsilon > 0$ and $0 < \widetilde{\varepsilon} < \varepsilon/(1+\varepsilon)$, then the decomposition algorithm using the generic form of optimality cut $\delta \geq q(\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}; \boldsymbol{c})$ in the master problem (12) terminates in a finite number of iterations.

Proof. Proof of Lemma 6. Note that Algorithm 3 proceeds to either the exploitation or exploration step in each iteration. Thus, to show its convergence, it suffices to show that there is a finite number of exploration and exploitation steps. We divide the proof into three steps.

Step 1. We show that there is a finite number of exploration steps when the set $\mathcal{C}$ is finite. Specifically, we show that the scenario set $\overline{\mathcal{C}}$ is enlarged every time when Algorithm 3 proceeds to the exploration step. Suppose, on the contrary, that $\overline{\boldsymbol{c}} \in \mathcal{C}$ obtained by solving the subproblem (13) at some iteration h in step 2 belongs to the current scenario set $\overline{\mathcal{C}}$, i.e., $\overline{\boldsymbol{c}} \in \overline{\mathcal{C}}$, but the algorithm proceeds to the exploration step. Let

$$m(\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \boldsymbol{x}, \boldsymbol{u}) := \sum_{i \in I} c_i^{\mathrm{f}} o_i + \frac{1-\theta}{N} \sum_{n=1}^{N} \left(\sum_{i \in I} \sum_{j \in J} t_{i,j} x_{i,j}^n + \sum_{j \in J} c_j^{\mathrm{u}} u_j^n \right) + \theta \left(\sum_{k \in [K]} \mu_k \rho_k + \sum_{k' \in [K']} \sigma_{k'} \gamma_{k'} \right)$$

be the objective function of the master problem (12) excluding the term $\theta\delta$. Since $\overline{\boldsymbol{c}} \in \overline{\mathcal{C}}$, we have

$$\begin{aligned} m(\boldsymbol{o}^h, \boldsymbol{x}^h, \boldsymbol{u}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h) + \theta \sup_{\boldsymbol{c} \in \mathcal{C}} q(\boldsymbol{o}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h; \boldsymbol{c}) &= m(\boldsymbol{o}^h, \boldsymbol{x}^h, \boldsymbol{u}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h) + \theta \sup_{\boldsymbol{c} \in \overline{\mathcal{C}}} q(\boldsymbol{o}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h; \boldsymbol{c}) \\ &\leq m(\boldsymbol{o}^h, \boldsymbol{x}^h, \boldsymbol{u}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h) + \theta \delta^h \leq U^h, \end{aligned}$$

where the first inequality follows from constraints (A.19) that $\delta^h \geq \sup_{\boldsymbol{c} \in \overline{\mathcal{C}}} q(\boldsymbol{o}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h; \boldsymbol{c})$, and the second inequality follows from the definition that U^h is an upper bound of the optimal value of the master problem. This implies that $\overline{U}$ updated in step 2 satisfies

$$\overline{U} - U^h \leq \left[m(\boldsymbol{o}^h, \boldsymbol{x}^h, \boldsymbol{u}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h) + \theta \sup_{\boldsymbol{c} \in \mathcal{C}} q(\boldsymbol{o}^h, \boldsymbol{\rho}^h, \boldsymbol{\gamma}^h; \boldsymbol{c}) \right] - U^h \leq 0.$$

As such, we have $(\overline{U} - U^h)/\overline{U} < 0 < \widetilde{\varepsilon}$ and thus, Algorithm 3 proceeds to the exploitation step, which contradicts with the assumption that it proceeds to the exploration step. Therefore, since a new scenario $\bar{\boldsymbol{c}} \in \mathcal{C}$ is identified in every exploration step and the set $\mathcal{C}$ is finite, there is a finite number of exploration steps.

Step 2. Next, we show that there is a finite number of exploration steps when the set $\mathcal{C}$ is not finite. Let T be the number of exploration steps, which may be infinite. Let $\{(\boldsymbol{o}^t, \boldsymbol{x}^t, \boldsymbol{u}^t, \boldsymbol{\rho}^t, \boldsymbol{\gamma}^t, \delta^t)\}_{t=1}^T$ be the set of the best feasible solutions obtained from solving the master problem (12) in step 1 of iterations $t \in [T]$ at which Algorithm 3 visits the exploration step. Denote U^t and $\overline{U}^t$ as the upper bound of the optimal value of the master problem obtained in step 1 and the upper bound on $\upsilon^\star(\theta)$ computed in step 2, respectively. Also, denote $\boldsymbol{c}^t \in \mathcal{C}$ as the scenario obtained by solving the subproblem (13). Our goal is to show that T is upper bounded and thus finite. To prove this, since Algorithm 3 proceeds to the exploration step at iteration $t \in [T]$, we have $(\overline{U}^t - U^t)/\overline{U}^t \geq \widetilde{\varepsilon}$ for all $t \in [T]$. Therefore,

$$\begin{aligned}
\theta\Big[\sup_{\boldsymbol{c}\in\mathcal{C}} q(\boldsymbol{o}^t, \boldsymbol{\rho}^t, \boldsymbol{\gamma}^t; \boldsymbol{c}) - \delta^t\Big] &= \Big[m(\boldsymbol{o}^t, \boldsymbol{x}^t, \boldsymbol{u}^t, \boldsymbol{\rho}^t, \boldsymbol{\gamma}^t) + \theta \sup_{\boldsymbol{c}\in\mathcal{C}} q(\boldsymbol{o}^t, \boldsymbol{\rho}^t, \boldsymbol{\gamma}^t; \boldsymbol{c})\Big] - \Big[m(\boldsymbol{o}^t, \boldsymbol{x}^t, \boldsymbol{u}^t, \boldsymbol{\rho}^t, \boldsymbol{\gamma}^t) + \theta\delta^t\Big] \\
&\geq \overline{U}^t - U^t && \text{(A.21a)} \\
&\geq \widetilde{\varepsilon}\overline{U}^t && \text{(A.21b)} \\
&\geq \widetilde{\varepsilon}\upsilon^\star(\theta). && \text{(A.21c)}
\end{aligned}$$

Here, (A.21a) follows from the definitions of $\overline{U}^t$ and U^t, (A.21b) follows from $(\overline{U}^t - U^t)/\overline{U}^t \geq \widetilde{\varepsilon}$, and (A.21c) follows from $\overline{U}^t$ is an upper bound on $\upsilon^\star(\theta)$. By definition of q', (A.21) implies that

$$q'(\boldsymbol{y}^t; \boldsymbol{c}^t) = q(\boldsymbol{o}^t, \boldsymbol{\rho}^t, \boldsymbol{\gamma}^t; \boldsymbol{c}^t) - \delta^t \geq \frac{\widetilde{\varepsilon}\upsilon^\star(\theta)}{\theta}. \tag{A.22}$$

Since the new scenario $\boldsymbol{c}^t \in \mathcal{C}$ identified in iteration t is added to the scenario set $\overline{\mathcal{C}}$ from iteration t, for any subsequent iteration $t' > t$, the master problem (12) includes the constraint $\delta \geq q(\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}; \boldsymbol{c}^t)$. Therefore, since $(\boldsymbol{o}^{t'}, \boldsymbol{\rho}^{t'}, \boldsymbol{\gamma}^{t'}, \delta^{t'})$ is a feasible solution to the master problem in iteration t', we have

$$q'(\boldsymbol{y}^{t'}; \boldsymbol{c}^t) = q(\boldsymbol{o}^{t'}, \boldsymbol{\rho}^{t'}, \boldsymbol{\gamma}^{t'}; \boldsymbol{c}^t) - \delta^{t'} \leq 0. \tag{A.23}$$

Combining the two inequalities (A.22) and (A.23), we obtain

$$q'(\boldsymbol{y}^t; \boldsymbol{c}^t) - q'(\boldsymbol{y}^{t'}; \boldsymbol{c}^t) \geq \frac{\widetilde{\varepsilon}\upsilon^\star(\theta)}{\theta} \tag{A.24}$$

for all $t' > t$. If follows from the Lipschitz continuity of q' and (A.24) that $L\|\boldsymbol{y}^t - \boldsymbol{y}^{t'}\| \geq \widetilde{\varepsilon}\upsilon^\star(\theta)/\theta$, which in turn implies $\|\boldsymbol{y}^t - \boldsymbol{y}^{t'}\| \geq \widetilde{\varepsilon}\upsilon^\star(\theta)/(\theta L)$ for all $t' > t$. Define $\mathbb{B}_t := \mathbb{B}\big(\boldsymbol{y}^t, \widetilde{\varepsilon}\upsilon^\star(\theta)/(2\theta L)\big)$ as the norm ball centered at $\boldsymbol{y}^t$ with radius $\widetilde{\varepsilon}\upsilon^\star(\theta)/(2\theta L)$. Note that by construction, any two balls in $\{\mathbb{B}_t\}_{t=1}^T$ are not overlapping. Therefore, the total volume of the norm balls in $\{\mathbb{B}_t\}_{t=1}^T$ is equal to $T\big[\,\mathrm{Vol}(\mathbb{B}(\mathbf{0}, 1)) \cdot (\widetilde{\varepsilon}\upsilon^\star(\theta)/(2\theta L))^n\big]$, where $\mathrm{Vol}(\mathbb{B}(\mathbf{0}, 1))$ is the volume of a unit norm ball of

dimension n and $n = |I| + K + K' + 1$ is the dimension of $\boldsymbol{y} = (\boldsymbol{o}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \delta)$. By the assumption that the optimal solutions to the master problems are contained in a compact set, $\{\boldsymbol{y}^t\}_{t=1}^T$ is contained in some compact set $\mathcal{Y}$. Since $\mathcal{Y}$ is compact, there exists finite $R > 0$ such that $\mathcal{Y} \subseteq \mathbb{B}(\mathbf{0}, R)$. Thus, we have $\{\mathbb{B}_t\}_{t=1}^T \subseteq \mathbb{B}\big(\mathbf{0}, R + \widetilde{\varepsilon}\upsilon^\star(\theta)/(2\theta L)\big)$. This implies that

$$T\left[\operatorname{Vol}(\mathbb{B}(\mathbf{0},1)) \cdot \left(\frac{\widetilde{\varepsilon}\upsilon^\star(\theta)}{2\theta L}\right)^n\right] \leq \operatorname{Vol}(\mathbb{B}(\mathbf{0},1))\left(R + \frac{\widetilde{\varepsilon}\upsilon^\star(\theta)}{2\theta L}\right)^n$$

and thus, we have

$$T \leq \left(\frac{2\theta L R}{\widetilde{\varepsilon}\upsilon^\star(\theta)} + 1\right)^n < \infty.$$

This shows that T is finite, i.e., there is finite number of exploration steps.

Step 3. Finally, we show that there is a finite number of exploitation steps. This part of proof is the same as Tsang et al. (2023), and we provide the proof here for completeness. Suppose, on the contrary, that the master problem (12) with a given scenario set $\widetilde{\mathcal{C}} \subseteq \mathcal{C}$ is solved infinitely many times. Note that this happens only when the algorithm does not proceed to the exploration step nor terminate. In other words, from the first time we solve the master problem (12) with the scenario set $\widetilde{\mathcal{C}} \subseteq \mathcal{C}$, Algorithm 3 visits the exploitation step forever. If follows from Proposition 2 of Tsang et al. (2023) that an upper bound on the actual relative gap $(\overline{U} - L^p)/\overline{U}$ in each iteration is given by $1/[(1-\widetilde{\varepsilon})(1-\varepsilon^p_{\text{MP}})] - 1$. Since the relative optimality gap of the master problem shrinks by a factor of $\alpha \in (0,1)$ every time when the algorithm proceeds to the exploitation step, we have $\varepsilon^p_{\text{MP}} \to 0$. This implies that the upper bound on the actual relative gap $1/[(1-\widetilde{\varepsilon})(1-\varepsilon^p_{\text{MP}}] - 1$ converges to $\widetilde{\varepsilon}/(1-\widetilde{\varepsilon})$, which is less than ε by our assumption. This shows that Algorithm 3 terminates after a sufficiently large number of iterations, which contradicts that the master problem (12) with the scenario set $\widetilde{\mathcal{C}} \subseteq \mathcal{C}$ is solved infinitely many times.

In the above three steps, we show that there is a finite number of exploration and exploitation steps. Hence, Algorithm 3 terminates in a finite number of iterations. □

To apply Lemma 6 to show the finite convergence of the four algorithms (i.e., the BD, pC&CG, dC&CG, and hC&CG algorithms), it suffices to show that the corresponding function q' in each of these algorithms is uniformly Lipschitz continuous over $\boldsymbol{c} \in \mathcal{C}$. Specifically, we want to show that there exists $L > 0$ such that for $\boldsymbol{y}_1 = (\boldsymbol{o}_1, \boldsymbol{\rho}_1, \boldsymbol{\gamma}_1, \delta_1)$ and $\boldsymbol{y}_2 = (\boldsymbol{o}_2, \boldsymbol{\rho}_2, \boldsymbol{\gamma}_2, \delta_2)$, we have $|q'(\boldsymbol{y}_1; \boldsymbol{c}) - q'(\boldsymbol{y}_2; \boldsymbol{c})| \leq L\|\boldsymbol{y}_1 - \boldsymbol{y}_2\|$ for all $\boldsymbol{c} \in \mathcal{C}$.

- In BD, for a given $\boldsymbol{c} = (\boldsymbol{d}, e) \in \mathcal{S} \times E = \mathcal{C}$, we have

$$q'(\boldsymbol{y}; \boldsymbol{c}) = q'_{\text{BD}}(\boldsymbol{y}; \boldsymbol{d}, e) = \sum_{i \in I} C_i o_i v_i^e + \sum_{j \in J} w_j^e d_j - \sum_{k \in [K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k' \in [K']} \gamma_{k'} g_{k'}(\boldsymbol{d}) - \delta,$$

  which is linear in $\boldsymbol{y}$. Therefore,

$$|q'(\boldsymbol{y}_1; \boldsymbol{c}) - q'(\boldsymbol{y}_2; \boldsymbol{c})|$$

$$
\begin{aligned}
&= \left| \sum_{i\in I} C_i(o_{1,i} - o_{2,i})v_i^e - \sum_{k\in[K]} (\rho_{1,k} - \rho_{2,k}) f_k(\boldsymbol{d}) - \sum_{k'\in[K']} (\gamma_{1,k'} - \gamma_{2,k'}) g_{k'}(\boldsymbol{d}) - (\delta_1 - \delta_2) \right| \\
&\le \sum_{i\in I} C_i |v_i^e| |o_{1,i} - o_{2,i}| + \sum_{k\in[K]} |f_k(\boldsymbol{d})| |\rho_{1,k} - \rho_{2,k}| + \sum_{k'\in[K']} |g_{k'}(\boldsymbol{d})| |\gamma_{1,k'} - \gamma_{2,k'}| + |\delta_1 - \delta_2| \qquad \text{(A.25)} \\
&\le \max\left\{ \max_{i\in I} \{C_i |v_i^e|\}, \max_{k\in[K]} |f_k(\boldsymbol{d})|, \max_{k'\in[K']} |g_{k'}(\boldsymbol{d})| \right\} \cdot \|\boldsymbol{y}_1 - \boldsymbol{y}_2\|_1 \\
&=: L(\boldsymbol{d}, e) \|\boldsymbol{y}_1 - \boldsymbol{y}_2\|_1, \qquad \text{(A.26)}
\end{aligned}
$$

where (A.25) follows from the triangular inequality. Note that since $\mathcal{S}$ is compact and E is finite, we can pick $L = \sup_{\boldsymbol{d}\in\mathcal{S}, e\in E} L(\boldsymbol{d}, e)$. This shows that $q'_{\mathrm{BD}}(\boldsymbol{y}; \boldsymbol{d}, e)$ is uniformly Lipschitz.

- In pC&CG, for a given $\boldsymbol{c} = \boldsymbol{d} \in \mathcal{S} = \mathcal{C}$, we have

$$
q'(\boldsymbol{y}; \boldsymbol{c}) = q'_{\mathrm{pC\&CG}}(\boldsymbol{y}; \boldsymbol{d}) = \max_{e\in E} \left\{ \sum_{i\in I} C_i o_i v_i^e + \sum_{j\in J} w_j^e d_j \right\} - \sum_{k\in[K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'} g_{k'}(\boldsymbol{d}) - \delta,
$$

which is linear in $(\boldsymbol{\rho}, \boldsymbol{\gamma}, \delta)$. Note that

$$
\begin{aligned}
|q'(\boldsymbol{y}_1; \boldsymbol{c}) - q'(\boldsymbol{y}_2; \boldsymbol{c})| \le &\left| \max_{e\in E} \left\{ \sum_{i\in I} C_i o_{1,i} v_i^e + \sum_{j\in J} w_j^e d_j \right\} - \max_{e\in E} \left\{ \sum_{i\in I} C_i o_{2,i} v_i^e + \sum_{j\in J} w_j^e d_j \right\} \right| \\
&+ \left| \sum_{k\in[K]} (\rho_{1,k} - \rho_{2,k}) f_k(\boldsymbol{d}) + \sum_{k'\in[K']} (\gamma_{1,k'} - \gamma_{2,k'}) g_{k'}(\boldsymbol{d}) + (\delta_1 - \delta_2) \right|. \qquad \text{(A.27)}
\end{aligned}
$$

For the first term on the right-hand-side of (A.27), we have

$$
\begin{aligned}
&\left| \max_{e\in E} \left\{ \sum_{i\in I} C_i o_{1,i} v_i^e + \sum_{j\in J} w_j^e d_j \right\} - \max_{e\in E} \left\{ \sum_{i\in I} C_i o_{2,i} v_i^e + \sum_{j\in J} w_j^e d_j \right\} \right| \\
&\le \max_{e\in E} \left\{ \sum_{i\in I} C_i |v_i^e| |o_{1,i} - o_{2,i}| \right\} \\
&\le \max_{e\in E} \max_{i\in I} \{C_i |v_i^e|\} \cdot \|\boldsymbol{o}_1 - \boldsymbol{o}_2\|_1.
\end{aligned}
$$

The second term on the right-hand-side of (A.27) is same as those in q'_{BD}. Thus, following a similar argument as in (A.26), we can show that $q'_{\mathrm{pC\&CG}}(\boldsymbol{y}; \boldsymbol{d})$ is uniformly Lipschitz.

- In dC&CG, for a given $\boldsymbol{c} = e \in E = \mathcal{C}$, we have

$$
q'(\boldsymbol{y}; \boldsymbol{c}) = q'_{\mathrm{dC\&CG}}(\boldsymbol{y}; e) = \sum_{i\in I} C_i o_i v_i^e + \max_{\boldsymbol{d}\in\mathcal{S}} \left\{ \sum_{j\in J} w_j^e d_j - \sum_{k\in[K]} \rho_k f_k(\boldsymbol{d}) - \sum_{k'\in[K']} \gamma_{k'} g_{k'}(\boldsymbol{d}) \right\} - \delta,
$$

which is linear in $(\boldsymbol{o}, \delta)$. Note that

$$
|q'(\boldsymbol{y}_1; \boldsymbol{c}) - q'(\boldsymbol{y}_2; \boldsymbol{c})| \le \left| \max_{\boldsymbol{d}\in\mathcal{S}} \left\{ (\boldsymbol{w}^e)^\top \boldsymbol{d} - \boldsymbol{\rho}_1^\top \boldsymbol{f}(\boldsymbol{d}) - \boldsymbol{\gamma}_1^\top \boldsymbol{g}(\boldsymbol{d}) \right\} - \max_{\boldsymbol{d}\in\mathcal{S}} \left\{ (\boldsymbol{w}^e)^\top \boldsymbol{d} - \boldsymbol{\rho}_2^\top \boldsymbol{f}(\boldsymbol{d}) - \boldsymbol{\gamma}_2^\top \boldsymbol{g}(\boldsymbol{d}) \right\} \right|
$$

$$+\left|\sum_{i\in I} C_i(o_{1,i}-o_{2,i})v_i^e-(\delta_1-\delta_2)\right|. \tag{A.28}$$

For the first term on the right-hand-side of (A.28), we have

$$\begin{aligned}
&\left|\max_{\boldsymbol{d}\in\mathcal{S}}\left\{(\boldsymbol{w}^e)^\top\boldsymbol{d}-\boldsymbol{\rho}_1^\top\boldsymbol{f}(\boldsymbol{d})-\boldsymbol{\gamma}_1^\top\boldsymbol{g}(\boldsymbol{d})\right\}-\max_{\boldsymbol{d}\in\mathcal{S}}\left\{(\boldsymbol{w}^e)^\top\boldsymbol{d}-\boldsymbol{\rho}_2^\top\boldsymbol{f}(\boldsymbol{d})-\boldsymbol{\gamma}_2^\top\boldsymbol{g}(\boldsymbol{d})\right\}\right| \\
&\leq \max_{\boldsymbol{d}\in\mathcal{S}}\left\{\|\boldsymbol{f}(\boldsymbol{d})\|_\infty\|\boldsymbol{\rho}_1-\boldsymbol{\rho}_2\|_1+\|\boldsymbol{g}(\boldsymbol{d})\|_\infty\|\boldsymbol{\gamma}_1-\boldsymbol{\gamma}_2\|_1\right\} \\
&\leq \max_{\boldsymbol{d}\in\mathcal{S}}\max\left\{\max_{k\in[K]}|f_k(\boldsymbol{d})|,\max_{k'\in[K']}|g_{k'}(\boldsymbol{d})|\right\}\cdot\|(\boldsymbol{\rho}_1,\boldsymbol{\gamma}_1)-(\boldsymbol{\rho}_2,\boldsymbol{\gamma}_2)\|_1.
\end{aligned}$$

The second term on the right-hand-side of (A.28) is same as those in q'_{BD}. Thus, following a similar argument in (A.26), we can show that $q'_{\mathrm{dC\&CG}}(\boldsymbol{y};\boldsymbol{d})$ is uniformly Lipschitz.

- In hC&CG, for a given $\boldsymbol{c}=(\boldsymbol{d},e)\in\mathcal{S}\times E=\mathcal{C}$, we have

$$q'(\boldsymbol{y};\boldsymbol{c})=q'_{\mathrm{hC\&CG}}(\boldsymbol{y};\boldsymbol{d},e)=\max\left\{q'_{\mathrm{pC\&CG}}(\boldsymbol{y};\boldsymbol{d}),\,q'_{\mathrm{dC\&CG}}(\boldsymbol{y};e)\right\},$$

  which is the maximum of two uniformly Lipschitz functions. Thus, this shows that $q'_{\mathrm{hC\&CG}}(\boldsymbol{y};\boldsymbol{d})$ is also uniformly Lipschitz.

Combining the four cases, we have that the hC&CG, pC&CG, dC&CG, and BD algorithms terminate in a finite number of iterations. □

Appendix B. Subroutines of the Spectrum Search Algorithm

Appendix B.1. A Gift Wrapping Algorithm for Finding Two-Dimensional Convex Hulls

In Algorithm 4, we present a gift-wrapping algorithm tailored to our context to obtain the extreme points of $\mathrm{conv}\left(\mathrm{hyp}_{[0,1]}(\breve{v})\right)\subseteq[0,1]\times\mathbb{R}$ (Jarvis, 1973). We initialize this algorithm with the set of extreme points $\mathcal{E}=\{(\theta_1,\underline{v}_1)\}$ and $\ell=1$, where ℓ is the index of the *rightmost* extreme point in the current set $\mathcal{E}$, i.e., the *largest* index such that $(\theta_\ell,\underline{v}_\ell)$ in $\mathcal{E}$ is an extreme point. In each iteration, we compute the slopes between the rightmost identified extreme point $(\theta_\ell,\underline{v}_\ell)$ and the points $\{(\theta_m,\underline{v}_m)\}_{m>\ell}$, i.e., points to the right of $(\theta_\ell,\underline{v}_\ell)$. We then identify the largest index $m_{\max}>\ell$ such that the slope between $(\theta_\ell,\underline{v}_\ell)$ and $(\theta_{m_{\max}},\underline{v}_{m_{\max}})$ is the greatest. This indicates that $(\theta_{m_{\max}},\underline{v}_{m_{\max}})$ is the next extreme point to the right of $(\theta_\ell,\underline{v}_\ell)$. Next, we update the lower bounding value $\underline{v}_m$ for any $m\in\{\ell+1,\ldots,m_{\max}-1\}$. Finally, we update the index ℓ to $m_{\max}$ and enlarge the set of extreme points $\mathcal{E}$ by $(\theta_\ell,\underline{v}_\ell)$. Note that the lower bounding function $\widehat{v}$ is the one linearly interpolating the points $\{(\theta_m,\underline{v}_m)\}_{m=1}^M$ updated using Algorithm 4.

Appendix B.2. Derivatives of the Lower and Upper Bounding Functions

In Algorithm 1, we need to evaluate the slope p_m of the lower bounding function $\widehat{v}$ on each interval $[\theta_{m-1},\theta_m]$ for $m\in\{2,\ldots,M\}$. We also need the left derivative $s_-(\theta)$ and the right derivative $s_+(\theta)$ of the upper bounding function $\widetilde{v}$ at any point $\theta\in(0,1)$. In this appendix, we provide details on how to obtain these elements.

Algorithm 4: A gift wrapping algorithm to construct the lower bounding function $\widehat{\upsilon}$ for a given set of points $\{(\theta_m, \underline{\upsilon}_m)\}_{m=1}^M$

Initialization: $\mathcal{E} \leftarrow \{(\theta_1, \underline{\upsilon}_1)\}$, $\ell \leftarrow 1$
while $\ell < M$ **do**
  Compute $\texttt{slope}_{m'} = (\underline{\upsilon}_{m'} - \underline{\upsilon}_\ell)/(\theta_{m'} - \theta_\ell)$ for all $m' > \ell$.
  Identify the index $m_{\max} = \max\{m' > \ell \mid \texttt{slope}_{m'} \geq \texttt{slope}_{m''}, \forall m'' > \ell\}$.
  If $m_{\max} > \ell + 1$, update the lower bounding values $\underline{\upsilon}_m = \underline{\upsilon}_\ell + \texttt{slope}_{m_{\max}}(\theta_m - \theta_\ell)$ for all $m \in \{\ell + 1, \ldots, m_{\max} - 1\}$.
  Set $\ell \leftarrow m_{\max}$ and update $\mathcal{E} \leftarrow \mathcal{E} \cup \{(\theta_\ell, \underline{\upsilon}_\ell)\}$.

Recall that the lower bounding function $\widehat{\upsilon}$ is the function linearly interpolating the points $\{(\theta_m, \underline{\upsilon}_m)\}_{m=1}^M$, which we update in step 2d of Algorithm 1 using Algorithm 4 (see Appendix B.1). For each $m \in \{2, \ldots, M\}$, we compute slope p_m of $\widehat{\upsilon}$ on interval $[\theta_{m-1}, \theta_m]$ as $p_m = (\underline{\upsilon}_m - \underline{\upsilon}_{m-1})/(\theta_m - \theta_{m-1})$.

To compute the left derivative $s_-(\theta)$ and right derivative $s_+(\theta)$ of the upper bounding function $\widetilde{\upsilon}$ at θ, we define

$$\mathcal{M}(\theta) = \arg\min_{m \in [M]} \left\{ \sum_{i \in I} c_i^{\mathrm{f}}(\widehat{\boldsymbol{o}}_m)_i + (1 - \theta)\widehat{\varphi}_m^{\mathrm{SP}} + \theta \varphi_m^{\mathrm{DRO}} \right\}.$$

Then, we compute $s_-(\theta) = \max_{m \in \mathcal{M}(\theta)} \left\{\varphi_m^{\mathrm{DRO}} - \widehat{\varphi}_m^{\mathrm{SP}}\right\}$ and $s_+(\theta) = \min_{m \in \mathcal{M}(\theta)} \left\{\varphi_m^{\mathrm{DRO}} - \widehat{\varphi}_m^{\mathrm{SP}}\right\}$.

Appendix C. Tailored Reformulations for Some Specific Ambiguity Sets

In this section, we provide tailored reformulations of the TRO-CFLP model and valid inequalities for two popular choices of ambiguity sets: the mean-support (MS) ambiguity set and the mean-absolute-deviation (MAD) ambiguity set. Recall that the MS ambiguity set is defined by letting $K = |J|$ with $f_k(\boldsymbol{d}) = d_k$ for $k \in J$ and $K' = 0$ in (4), whereas the MAD ambiguity set is defined by letting $K = |J|$ with $f_k(\boldsymbol{d}) = d_k$ for $k \in J$ and $K' = |J|$ with $g_{k'}(\boldsymbol{d}) = |d_{k'} - \mu_{k'}|$ for $k' \in J$ in (4).

In Appendix C.1, we derive the equivalent reformulations of the TRO-CFLP model (2) using the results of Proposition 1. In Appendix C.2, we derive linear reformulations of the optimality cut (12e). In Section Appendix C.3, we derive equivalent MILP reformulations of the associated subproblem (13). Finally, in Appendix C.4, we derive tailored valid inequalities of the TRO-CFLP model using the MS and MAD ambiguity sets. All the proofs in this section are provided in Appendix A.

Appendix C.1. Reformulations of the TRO-CFLP Model

Using the results of Proposition 1, in Corollary 1, we provide equivalent reformulations of the TRO-CFLP model (2) using the MS and MAD ambiguity sets.

Corollary 1. *Consider the TRO-CFLP problem in (5).*

(a) Suppose that $K = |J|$ and $f_k(\boldsymbol{d}) = d_k$ for $k \in J$ and $K' = 0$ (as in the MS ambiguity set). In

this case, problem (5) is equivalent to

$$\underset{\boldsymbol{o}\in\mathcal{O},\boldsymbol{x},\boldsymbol{u},\boldsymbol{\rho},\delta}{\text{minimize}} \quad \sum_{i\in I} c_i^{\mathrm{f}} o_i + \frac{1-\theta}{N}\sum_{n=1}^{N}\left(\sum_{i\in I}\sum_{j\in J} t_{i,j}x_{i,j}^n + \sum_{j\in J} c_j^{\mathrm{u}} u_j^n\right) + \theta\left(\boldsymbol{\mu}^\top\boldsymbol{\rho} + \delta\right) \tag{C.1a}$$

$$\text{subject to} \quad \sum_{j\in J} x_{i,j}^n \le C_i o_i, \quad \forall i\in I,\, n\in[N], \tag{C.1b}$$

$$\sum_{i\in I} x_{i,j}^n + u_j \ge \widehat{d}_j^n, \quad \forall j\in J,\, n\in[N], \tag{C.1c}$$

$$\delta \ge Q(\boldsymbol{o},\boldsymbol{d}) - \sum_{j\in J}\rho_j d_j, \quad \forall \boldsymbol{d}\in\mathcal{S}. \tag{C.1d}$$

(b) Suppose that $K=|J|$, $f_k(\boldsymbol{d})=d_k$ for $k\in J$ and $K'=|J|$, and $g_{k'}(\boldsymbol{d})=|d_{k'}-\mu_{k'}|$ for $k'\in J$ (as in the MAD ambiguity set). In this case, problem (5) is equivalent to

$$\underset{\boldsymbol{o}\in\mathcal{O},\boldsymbol{x},\boldsymbol{u},\boldsymbol{\rho},\boldsymbol{\gamma}\ge\boldsymbol{0},\delta}{\text{minimize}} \quad \sum_{i\in I} c_i^{\mathrm{f}} o_i + \frac{1-\theta}{N}\sum_{n=1}^{N}\left(\sum_{i\in I}\sum_{j\in J} t_{i,j}x_{i,j}^n + \sum_{j\in J} c_j^{\mathrm{u}} u_j^n\right) + \theta\left(\boldsymbol{\mu}^\top\boldsymbol{\rho} + \boldsymbol{\sigma}^\top\boldsymbol{\gamma} + \delta\right) \tag{C.2a}$$

$$\text{subject to} \quad \sum_{j\in J} x_{i,j}^n \le C_i o_i, \quad \forall i\in I,\, n\in[N], \tag{C.2b}$$

$$\sum_{i\in I} x_{i,j}^n + u_j \ge \widehat{d}_j^n, \quad \forall j\in J,\, n\in[N], \tag{C.2c}$$

$$\delta \ge Q(\boldsymbol{o},\boldsymbol{d}) - \sum_{j\in J}\rho_j d_j - \sum_{j\in J}\gamma_j|d_j-\mu_j|, \quad \forall \boldsymbol{d}\in\mathcal{S}. \tag{C.2d}$$

Appendix C.2. Linear Reformulations of the Optimality Cut (12e)

In Proposition 8, we show that if the TRO set is defined using the MS and MAD ambiguity sets, the optimality cut (12e) is equivalent to a set of linear constraints ((C.3) and (C.4), respectively). Thus, in such cases, the master problem (12) in Algorithm 3 is linear.

Proposition 8. *Consider optimality cut (12e).*

(a) Suppose that $K=|J|$ and $f_k(\boldsymbol{d})=d_k$ for $k\in J$ and $K'=0$ (as in the MS ambiguity set), then optimality cut (12e) is equivalent to

$$\delta \ge \sum_{i\in I} C_i v_i^e o_i + \sum_{j\in J}\left(\overline{\pi}_j^e\overline{d}_j - \underline{\pi}_j^e\underline{d}_j\right), \tag{C.3a}$$

$$\underline{\pi}_j^e - \overline{\pi}_j^e = \rho_j - w_j^e, \quad \forall j\in J, \tag{C.3b}$$

$$\underline{\pi}_j^e \ge 0,\ \overline{\pi}_j^e \ge 0, \quad \forall j\in J. \tag{C.3c}$$

(b) Suppose that $K=|J|$, $f_k(\boldsymbol{d})=d_k$ for $k\in J$ and $K'=|J|$, and $g_{k'}(\boldsymbol{d})=|d_{k'}-\mu_{k'}|$ for $k'\in J$ (as in the MAD ambiguity set), then optimality cut (12e) is equivalent to

$$\delta \ge \sum_{i\in I} C_i v_i^e o_i + \sum_{j\in J}\left(\overline{\pi}_j^e\overline{d}_j - \underline{\pi}_j^e\underline{d}_j + \overline{\theta}_j^e\mu_j - \underline{\theta}_j^e\mu_j\right), \tag{C.4a}$$

$$\underline{\pi}_j^e - \overline{\pi}_j^e + \underline{\theta}_j^e - \overline{\theta}_j^e = \rho_j - w_j^e, \quad \forall j \in J, \tag{C.4b}$$

$$\overline{\theta}_j^e + \underline{\theta}_j^e = \gamma_j, \quad \forall j \in J, \tag{C.4c}$$

$$\underline{\pi}_j^e \geq 0,\ \overline{\pi}_j^e \geq 0,\ \underline{\theta}_j^e \geq 0,\ \overline{\theta}_j^e \geq 0, \quad \forall j \in J. \tag{C.4d}$$

Proof. Proof. We first prove part (a). Note that the inner supremum in (12e) is equivalent to

$$\underset{\boldsymbol{d}}{\text{maximize}} \quad \sum_{j \in J} w_j^e d_j - \sum_{j \in J} \rho_j d_j \tag{C.5a}$$

$$\text{subject to} \quad \underline{d}_j \leq d_j \leq \overline{d}_j, \quad \forall j \in J. \tag{C.5b}$$

By LP duality, (C.5) is equivalent to

$$\underset{\boldsymbol{\pi}^e}{\text{minimize}} \quad \sum_{j \in J} \left(\overline{\pi}_j^e \overline{d}_j - \underline{\pi}_j^e \underline{d}_j \right) \tag{C.6a}$$

$$\text{subject to} \quad \underline{\pi}_j^e - \overline{\pi}_j^e = \rho_j - w_j^e, \quad \forall j \in J, \tag{C.6b}$$

$$\underline{\pi}_j^e \geq 0,\ \overline{\pi}_j^e \geq 0, \quad \forall j \in J. \tag{C.6c}$$

Let $\mathcal{D} = \{\boldsymbol{\pi}^e \mid \text{(C.6b)–(C.6c)}\}$ be the dual feasible set. We can rewrite the optimality cut (12e) as

$$\delta \geq \sum_{i \in I} C_i v_i^e o_i + \min_{\boldsymbol{\pi}^e \in \mathcal{D}} \left\{ \sum_{j \in J} \left(\overline{\pi}_j^e \overline{d}_j - \underline{\pi}_j^e \underline{d}_j \right) \right\},$$

which is equivalent to (C.3).

Next, we prove part (b). Note that the inner supremum in (12e) is equivalent to

$$\underset{\boldsymbol{d}}{\text{maximize}} \quad \sum_{j \in J} w_j^e d_j - \sum_{j \in J} \rho_j d_j - \sum_{j \in J} \gamma_j |d_j - \mu_j| \tag{C.7a}$$

$$\text{subject to} \quad \underline{d}_j \leq d_j \leq \overline{d}_j, \quad \forall j \in J. \tag{C.7b}$$

Introducing the auxiliary variable η_j to linearize the absolute value $|d_j - \mu_j|$ for all $j \in J$, we can reformulate (C.7) into

$$\underset{\boldsymbol{d}, \boldsymbol{\eta}}{\text{maximize}} \quad \sum_{j \in J} w_j^e d_j - \sum_{j \in J} \rho_j d_j - \sum_{j \in J} \gamma_j \eta_j \tag{C.8a}$$

$$\text{subject to} \quad \underline{d}_j \leq d_j \leq \overline{d}_j, \quad \forall j \in J, \tag{C.8b}$$

$$\eta_j \geq d_j - \mu_j,\ \eta_j \geq \mu_j - d_j, \quad \forall j \in J. \tag{C.8c}$$

By LP duality, (C.8) is equivalent to

$$\underset{\boldsymbol{\pi}^e, \boldsymbol{\theta}^e}{\text{minimize}} \quad \sum_{j \in J} \left(\overline{\pi}_j^e \overline{d}_j - \underline{\pi}_j^e \underline{d}_j + \overline{\theta}_j^e \mu_j - \underline{\theta}_j^e \mu_j \right) \tag{C.9a}$$

$$\text{subject to} \quad \underline{\pi}_j^e - \overline{\pi}_j^e + \underline{\theta}_j^e - \overline{\theta}_j^e = \rho_j - w_j^e, \quad \forall j \in J, \tag{C.9b}$$

$$\overline{\theta}_j^e + \underline{\theta}_j^e = \gamma_j, \quad \forall j \in J, \tag{C.9c}$$

$$\underline{\pi}_j^e \geq 0,\ \overline{\pi}_j^e \geq 0,\ \underline{\theta}_j^e \geq 0,\ \overline{\theta}_j^e \geq 0, \quad \forall j \in J. \tag{C.9d}$$

Let $\mathcal{D} = \{(\boldsymbol{\pi}^e, \boldsymbol{\theta}^e) \mid \text{(C.9b)–(C.9d)}\}$ be the dual feasible set. We can rewrite the optimality cut (12e) as

$$\delta \geq \sum_{i \in I} C_i v_i^e o_i + \min_{(\boldsymbol{\pi}^e, \boldsymbol{\theta}^e) \in \mathcal{D}} \left\{ \sum_{j \in J} \left(\overline{\pi}_j^e \overline{d}_j - \underline{\pi}_j^e \underline{d}_j + \overline{\theta}_j^e \mu_j - \underline{\theta}_j^e \mu_j \right) \right\},$$

which is equivalent to (C.4). □

Appendix C.3. MILP Reformulations of the Subproblem (13)

In Proposition 9, we derive equivalent MILP reformulations of the subproblem (13) in Algorithm 3 for the MS and MAD ambiguity sets. Recall that $\Pi = \{(\boldsymbol{w}, \boldsymbol{v}) \mid w_j + v_i \leq t_{i,j},\ 0 \leq w_j \leq c_j^{\mathrm{u}},\ v_i \leq 0,\ \forall i \in I,\ j \in J\}$ is the dual feasible set of the second-stage problem.

Proposition 9. *Consider subproblem (13).*

(a) Suppose that $K = |J|$ and $f_k(\boldsymbol{d}) = d_k$ for $k \in J$ and $K' = 0$ (as in the MS ambiguity set). In this case, subproblem (13) is equivalent to the following MILP.

$$\underset{\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{a}, \boldsymbol{\beta}}{\text{maximize}} \quad \sum_{i \in I} C_i o_i^h v_i + \sum_{j \in J} \left[\underline{d}_j (w_j - \rho_j^h) + \Delta d_j (\beta_j - a_j \rho_j^h) \right] \tag{C.10a}$$

$$\text{subject to} \quad (\boldsymbol{w}, \boldsymbol{v}) \in \Pi, \tag{C.10b}$$

$$\beta_j \geq 0,\ \beta_j \geq w_j + c_j^{\mathrm{u}}(a_j - 1),\ \beta_j \leq w_j,\ \beta_j \leq c_j^{\mathrm{u}} a_j, \quad \forall j \in J, \tag{C.10c}$$

$$a_j \in \{0, 1\}, \quad \forall j \in J, \tag{C.10d}$$

where $\Delta d_j = \overline{d}_j - \underline{d}_j$ for $j \in J$.

(b) Suppose that $K = |J|$, $f_k(\boldsymbol{d}) = d_k$ for $k \in J$ and $K' = |J|$, and $g_{k'}(\boldsymbol{d}) = |d_{k'} - \mu_{k'}|$ for $k' \in J$ (as in the MAD ambiguity set). In this case, subproblem (13) is equivalent to the following MILP.

$$\underset{\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\iota}}{\text{maximize}} \quad \sum_{i \in I} C_i o_i^h v_i + \sum_{j \in J} \left[\mu_j (w_j - \rho_j) - \underline{\Delta}_j (\beta_j - a_j \rho_j^h + a_j \gamma_j^h) + \overline{\Delta}_j (\iota_j - b_j \rho_j^h - b_j \gamma_j^h) \right] \tag{C.11a}$$

$$\text{subject to} \quad (\boldsymbol{w}, \boldsymbol{v}) \in \Pi, \tag{C.11b}$$

$$\beta_j \geq 0,\ \beta_j \geq w_j + c_j^{\mathrm{u}}(a_j - 1),\ \beta_j \leq w_j,\ \beta_j \leq c_j^{\mathrm{u}} a_j, \quad \forall j \in J, \tag{C.11c}$$

$$\iota_j \geq 0,\ \iota_j \geq w_j + c_j^{\mathrm{u}}(b_j - 1),\ \iota_j \leq w_j,\ \iota_j \leq c_j^{\mathrm{u}} b_j, \quad \forall j \in J, \tag{C.11d}$$

$$a_j + b_j \leq 1, \quad \forall j \in J, \tag{C.11e}$$

$$a_j \in \{0, 1\},\ b_j \in \{0, 1\}, \quad \forall j \in J, \tag{C.11f}$$

where $\overline{\Delta}_j = \overline{d}_j - \mu_j$ and $\underline{\Delta}_j = \mu_j - \underline{d}_j$ for $j \in J$.

Proof. Proof. First, we prove part (a). For MS ambiguity set, the corresponding subproblem (13) is

$$\underset{\boldsymbol{w},\boldsymbol{v},\boldsymbol{d}}{\text{maximize}} \quad \sum_{i\in I} C_i o_i^h v_i + \sum_{j\in J} d_j(w_j - \rho_j^h) \tag{C.12a}$$

$$\text{subject to} \quad (\boldsymbol{w},\boldsymbol{v}) \in \Pi, \tag{C.12b}$$

$$\underline{d}_j \le d_j \le \overline{d}_j, \quad \forall j \in J. \tag{C.12c}$$

Problem (C.12) is not linear due to the bilinear terms $d_j w_j$ in the objective function. To linearize the bilinear terms, we first observe that we can optimize the second term in (C.12a) over $\boldsymbol{d}$, i.e.,

$$\underset{\boldsymbol{d}}{\text{maximize}} \quad \sum_{j\in J} d_j(w_j - \alpha_j) \tag{C.13a}$$

$$\text{subject to} \quad \underline{d}_j \le d_j \le \overline{d}_j, \quad \forall j \in J. \tag{C.13b}$$

Objective (C.13a) is separable in j and is linear in d_j. It follows that the optimal solution to (C.13) is attained at either the lower bound ($d_j = \underline{d}_j$) or upper bound ($d_j = \overline{d}_j$). Thus, we can reformulate (C.13) as a binary integer program as follows. We first define a binary variable a_j that equals 1 if $d_j = \overline{d}_j$ and is zero if $d_j = \underline{d}_j$ for $j \in J$. Then, we replace d_j with $\underline{d}_j + a_j \Delta d_j$ in the objective, where $\Delta d_j = \overline{d}_j - \underline{d}_j$ for $j \in J$. The resulting equivalent reformulation of (C.13) is as follows.

$$\underset{\boldsymbol{a}}{\text{maximize}} \quad \sum_{j\in J} \Big[\underline{d}_j(w_j - \alpha_j) + a_j \Delta d_j (w_j - \alpha_j)\Big] \tag{C.14a}$$

$$\text{subject to} \quad a_j \in \{0,1\}, \quad \forall j \in J. \tag{C.14b}$$

Using (C.14), we derive the following equivalent reformulation of (C.12):

$$\underset{\boldsymbol{w},\boldsymbol{v},\boldsymbol{a}}{\text{maximize}} \quad \sum_{i\in I} C_i o_i^h v_i + \sum_{j\in J} \Big[\underline{d}_j(w_j - \alpha_j) + a_j \Delta d_j (w_j - \alpha_j)\Big] \tag{C.15a}$$

$$\text{subject to} \quad (\boldsymbol{w},\boldsymbol{v}) \in \Pi, \tag{C.15b}$$

$$a_j \in \{0,1\}, \quad \forall j \in J. \tag{C.15c}$$

Finally, note that (C.15) is non-linear due to the bilinear terms $a_j w_j$, which is a product of a binary variable a_j and a bounded variable w_j. Therefore, to reformulate (C.15) into an equivalent MILP, we introduce variables $\beta_j = a_j w_j$ and inequalities (C.10c). This gives the desired MILP reformulation (C.10).

Next, we prove part (b). For MS ambiguity set, the corresponding subproblem (13) is

$$\underset{\boldsymbol{w},\boldsymbol{v},\boldsymbol{d}}{\text{maximize}} \quad \sum_{i\in I} C_i o_i^h v_i + \sum_{j\in J} \Big[d_j(w_j - \rho_j^h) - |d_j - \mu_j|\gamma_j^h\Big] \tag{C.16a}$$

$$\text{subject to} \quad (\boldsymbol{w},\boldsymbol{v}) \in \Pi, \tag{C.16b}$$

$$\underline{d}_j \leq d_j \leq \overline{d}_j, \quad \forall j \in J. \tag{C.16c}$$

It is easy to verify that in the optimal solution to (C.16) $d_j^\star$ satisfies $d_j^\star \in \{\underline{d}_j, \mu, \overline{d}_j\}$. To reformulate the non-linear objective (C.16a), let a_j and q_j be two binary variables and define $d_i = \mu_j - a_j\underline{\Delta}_j + b_j\overline{\Delta}_j$, where $\overline{\Delta}_j = \overline{d}_j - \mu_j$ and $\underline{\Delta}_j = \mu_j - \underline{d}_j$ for $j \in J$. Then, problem (C.16) is equivalent to

$$\begin{aligned}
\underset{\boldsymbol{w},\boldsymbol{v},\boldsymbol{a},\boldsymbol{b}}{\text{maximize}} \quad & \sum_{i\in I} C_i o_i^h v_i + \sum_{j\in J}\Big[(\mu_j - a_j\underline{\Delta}_j + b_j\overline{\Delta}_j)(w_j - \rho_j^h) - (a_j\underline{\Delta}_j + b_j\overline{\Delta}_j)\gamma_j^h\Big] && \text{(C.17a)}\\
\text{subject to} \quad & (\boldsymbol{w},\boldsymbol{v}) \in \Pi, && \text{(C.17b)}\\
& a_j + b_j \leq 1, \quad \forall j \in J, && \text{(C.17c)}\\
& a_j \in \{0,1\},\ b_j \in \{0,1\}, \quad \forall j \in J. && \text{(C.17d)}
\end{aligned}$$

Finally, note that (C.17) is non-linear due to the bilinear terms $a_j w_j$ (resp. $b_j w_j$), which is a product of a binary variable a_j (resp. b_j) and a bounded variable w_j. Therefore, to reformulate (C.17), we define $\beta_j = a_j w_j$ and $\iota_j = b_j w_j$ for $j \in J$, and introduce inequalities (C.11c)–(C.11d). This gives the desired MILP reformulation (C.11). □

Appendix C.4. Additional Valid Inequalities

In this section, we derive valid inequalities related to the dual variables $(\boldsymbol{\rho},\boldsymbol{\gamma})$ in the reformulation (5). First, in Proposition 10, we derive valid inequalities for the reformulation (C.1) under the MS ambiguity set.

Proposition 10. *The following inequalities are valid for* (C.1)*:*

$$0 \leq \rho_j \leq c_j^{\mathrm{u}}, \quad \forall j \in J. \tag{C.18}$$

Proof. Proof. Consider the reformulation of the DRO problem $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})]$ when $\mathcal{F}$ is the MS ambiguity set (see the proof of Proposition 1):

$$\underset{\boldsymbol{\rho}}{\text{minimize}} \quad \boldsymbol{\mu}^\top\boldsymbol{\rho} + \sup_{\boldsymbol{d}\in\mathcal{S}}\left\{Q(\boldsymbol{o},\boldsymbol{d}) - \sum_{j\in J}\rho_j d_j\right\}. \tag{C.19}$$

Using the dual reformulation of $Q(\boldsymbol{o},\boldsymbol{d})$, we can reformulate (C.19) into

$$\underset{\boldsymbol{\rho}}{\text{minimize}} \quad h(\boldsymbol{\rho};\boldsymbol{o}) := \boldsymbol{\mu}^\top\boldsymbol{\rho} + \sup_{(\boldsymbol{w},\boldsymbol{v})\in\Pi,\,\boldsymbol{d}\in\mathcal{S}}\left\{\sum_{i\in I} C_i o_i v_i + \sum_{j\in J}(w_j - \rho_j)d_j\right\}. \tag{C.20}$$

We first derive the upper bound on $\rho_{j'}$ for a fixed $j' \in J$. Note that for any $\rho_{j'} \geq c_{j'}^{\mathrm{u}}$, an optimal solution $\boldsymbol{d}_{j'}^\star$ to the inner supremum problem in (C.20) is $\boldsymbol{d}_{j'}^\star = \underline{d}_{j'}$. Indeed, if $\rho_{j'} \geq c_{j'}^{\mathrm{u}}$, then it follows from the bound on w_j (i.e., $0 \leq w_j \leq c_j^{\mathrm{u}}$) that $w_{j'} - \rho_{j'} \leq c_{j'}^{\mathrm{u}} - \rho_{j'} \leq 0$. Since $\underline{d}_{j'} \leq d_{j'} \leq \overline{d}_{j'}$, an optimal solution $\boldsymbol{d}_{j'}^\star$ to the inner supremum problem in (C.20) is $\boldsymbol{d}_{j'}^\star = \underline{d}_{j'}$. To show the upper bound on $\rho_{j'}$, for any fixed $\boldsymbol{o} \in \mathcal{O}$, let $\boldsymbol{\rho}^\star$ be an optimal solution to (C.20) with

$\rho^\star_{j'} > c^{\mathrm{u}}_{j'}$. Consider another feasible solution $\widetilde{\boldsymbol{\rho}}$ defined as follows: $\widetilde{\rho}_j = \rho^\star_j$ for $j \in J \setminus \{j'\}$ and $\widetilde{\rho}_{j'} = \rho^\star_j - \varepsilon \geq c^{\mathrm{u}}_{j'}$ with $\varepsilon \in (0, \rho^\star_{j'} - c^{\mathrm{u}}_{j'}]$. Note that

$$
\begin{aligned}
& h(\widetilde{\boldsymbol{\rho}}; \boldsymbol{o}) \\
&= \boldsymbol{\mu}^\top \widetilde{\boldsymbol{\rho}} + \sup_{(\boldsymbol{w},\boldsymbol{v}) \in \Pi,\, \boldsymbol{d} \in \mathcal{S}} \left\{ \sum_{i \in I} C_i o_i v_i + \sum_{j \in J} (w_j - \widetilde{\rho}_j) d_j \right\} \\
&= \boldsymbol{\mu}^\top \boldsymbol{\rho}^\star - \varepsilon \mu_{j'} + \sup_{(\boldsymbol{w},\boldsymbol{v}) \in \Pi,\, \boldsymbol{d} \in \mathcal{S}} \left\{ \sum_{i \in I} C_i o_i v_i + \sum_{j \in J \setminus \{j'\}} (w_j - \rho^\star_j) d_j + (w_{j'} - \rho^\star_{j'} + \varepsilon) d_{j'} \right\} && \text{(C.21a)} \\
&= \boldsymbol{\mu}^\top \boldsymbol{\rho}^\star - \varepsilon \mu_{j'} + \sup_{(\boldsymbol{w},\boldsymbol{v}) \in \Pi,\, \boldsymbol{d} \in \mathcal{S}} \left\{ \sum_{i \in I} C_i o_i v_i + \sum_{j \in J \setminus \{j'\}} (w_j - \rho^\star_j) d_j \right\} + (w_{j'} - \rho^\star_{j'} + \varepsilon) \underline{d}_{j'} && \text{(C.21b)} \\
&= \boldsymbol{\mu}^\top \boldsymbol{\rho}^\star - \varepsilon \mu_{j'} + \sup_{(\boldsymbol{w},\boldsymbol{v}) \in \Pi,\, \boldsymbol{d} \in \mathcal{S}} \left\{ \sum_{i \in I} C_i o_i v_i + \sum_{j \in J} (w_j - \rho^\star_j) d_j \right\} + \varepsilon \underline{d}_{j'} && \text{(C.21c)} \\
&= h(\boldsymbol{\rho}^\star; \boldsymbol{o}) + \varepsilon (\underline{d}_{j'} - \mu_{j'}) \\
&\leq h(\boldsymbol{\rho}^\star; \boldsymbol{o}). && \text{(C.21d)}
\end{aligned}
$$

Here, (C.21a) follows from the definition of $\widetilde{\boldsymbol{\rho}}$, (C.21b)–(C.21c) follow from the fact that $\underline{d}_{j'}$ is an optimal solution to the supremum problem over $d_{j'}$, (C.21d) follows from $\underline{d}_{j'} \leq \mu_{j'}$. Thus, the new feasible solution $\widetilde{\boldsymbol{\rho}}$ yields an optimal value no greater than $\boldsymbol{\rho}^\star$. Thus, without loss of optimality, one can impose the constraint $\rho_{j'} \leq c^{\mathrm{u}}_{j'}$.

We now derive the lower bound on $\rho_{j'}$ for a fixed $j' \in J$. Note that for any $\rho_{j'} \leq 0$, an optimal solution $\boldsymbol{d}^\star_{j'}$ to the inner supremum problem in (C.20) is $\boldsymbol{d}^\star_{j'} = \overline{d}_{j'}$. Indeed, if $\rho_{j'} \leq 0$, then it follows from the bound on w_j (i.e., $0 \leq w_j \leq c^{\mathrm{u}}_j$) that $w_{j'} - \rho_{j'} \geq 0$. Since $\underline{d}_{j'} \leq d_{j'} \leq \overline{d}_{j'}$, an optimal solution $\boldsymbol{d}^\star_{j'}$ to the inner supremum problem in (C.20) is $\boldsymbol{d}^\star_{j'} = \overline{d}_{j'}$. To show the lower bound on $\rho_{j'}$, for any fixed $\boldsymbol{o} \in \mathcal{O}$, let $\boldsymbol{\rho}^\star$ be an optimal solution to (C.20) with $\rho^\star_{j'} < 0$. Consider another feasible solution $\widetilde{\boldsymbol{\rho}}$ defined as follows: $\widetilde{\rho}_j = \rho^\star_j$ for $j \in J \setminus \{j'\}$ and $\widetilde{\rho}_{j'} = \rho^\star_j + \varepsilon \leq 0$ with $\varepsilon \in (0, -\rho^\star_{j'}]$. Note that

$$
\begin{aligned}
& h(\widetilde{\boldsymbol{\rho}}; \boldsymbol{o}) \\
&= \boldsymbol{\mu}^\top \widetilde{\boldsymbol{\rho}} + \sup_{(\boldsymbol{w},\boldsymbol{v}) \in \Pi,\, \boldsymbol{d} \in \mathcal{S}} \left\{ \sum_{i \in I} C_i o_i v_i + \sum_{j \in J} (w_j - \widetilde{\rho}_j) d_j \right\} \\
&= \boldsymbol{\mu}^\top \boldsymbol{\rho}^\star + \varepsilon \mu_{j'} + \sup_{(\boldsymbol{w},\boldsymbol{v}) \in \Pi,\, \boldsymbol{d} \in \mathcal{S}} \left\{ \sum_{i \in I} C_i o_i v_i + \sum_{j \in J \setminus \{j'\}} (w_j - \rho^\star_j) d_j + (w_{j'} - \rho^\star_{j'} - \varepsilon) d_{j'} \right\} && \text{(C.22a)} \\
&= \boldsymbol{\mu}^\top \boldsymbol{\rho}^\star + \varepsilon \mu_{j'} + \sup_{(\boldsymbol{w},\boldsymbol{v}) \in \Pi,\, \boldsymbol{d} \in \mathcal{S}} \left\{ \sum_{i \in I} C_i o_i v_i + \sum_{j \in J \setminus \{j'\}} (w_j - \rho^\star_j) d_j \right\} + (w_{j'} - \rho^\star_{j'} - \varepsilon) \overline{d}_{j'} && \text{(C.22b)} \\
&= \boldsymbol{\mu}^\top \boldsymbol{\rho}^\star + \varepsilon \mu_{j'} + \sup_{(\boldsymbol{w},\boldsymbol{v}) \in \Pi,\, \boldsymbol{d} \in \mathcal{S}} \left\{ \sum_{i \in I} C_i o_i v_i + \sum_{j \in J} (w_j - \rho^\star_j) d_j \right\} - \varepsilon \underline{d}_{j'} && \text{(C.22c)} \\
&= h(\boldsymbol{\rho}^\star; \boldsymbol{o}) + \varepsilon (\mu_{j'} - \overline{d}_{j'})
\end{aligned}
$$

$$\leq h(\boldsymbol{\rho}^\star;\boldsymbol{o}). \tag{C.22d}$$

Here, (C.22a) follows from the definition of $\widetilde{\boldsymbol{\rho}}$, (C.22b)–(C.22c) follow from the fact that $\overline{d}_{j'}$ is an optimal solution to the supremum problem over $d_{j'}$, (C.22d) follows from $\overline{d}_{j'} \geq \mu_{j'}$. Thus, the new feasible solution $\widetilde{\boldsymbol{\rho}}$ yields an optimal value no greater than $\boldsymbol{\rho}^\star$. Thus, without loss of optimality, one can impose the constraint $\rho_{j'} \geq 0$. □

Next, in Proposition 11, we derive valid inequalities for the reformulation (C.2) under the MAD ambiguity set.

Proposition 11. *The following inequalities are valid for* (C.2)*:*

$$-\gamma_j \leq \rho_j \leq c_j^{\mathrm{u}} + \gamma_j, \quad \forall j \in J. \tag{C.23}$$

Proof. Proof. Consider the reformulation of the DRO problem $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(\boldsymbol{o},\boldsymbol{d})]$ when $\mathcal{F}$ is the MAD ambiguity set (see the proof of Proposition 1):

$$\underset{\boldsymbol{\rho},\boldsymbol{\gamma}}{\text{minimize}} \quad \boldsymbol{\mu}^\top\boldsymbol{\rho} + \boldsymbol{\sigma}^\top\boldsymbol{\gamma} + \sup_{\boldsymbol{d}\in\mathcal{S}} \left\{ Q(\boldsymbol{o},\boldsymbol{d}) - \sum_{j\in J} \rho_j d_j - \sum_{j\in J} \gamma_j |d_j - \mu_j| \right\}. \tag{C.24}$$

Using the dual reformulation of $Q(\boldsymbol{o},\boldsymbol{d})$, we can reformulate (C.24) into

$$\underset{\boldsymbol{\rho},\boldsymbol{\gamma}}{\text{minimize}} \quad h(\boldsymbol{\rho},\boldsymbol{\gamma};\boldsymbol{o}) := \boldsymbol{\mu}^\top\boldsymbol{\rho} + \boldsymbol{\sigma}^\top\boldsymbol{\gamma} + \sup_{(\boldsymbol{w},\boldsymbol{v})\in\Pi,\, \boldsymbol{d}\in\mathcal{S}} \left\{ \sum_{i\in I} C_i o_i v_i + \sum_{j\in J} \left[(w_j - \rho_j) d_j + \gamma_j |d_j - \mu_j| \right] \right\}. \tag{C.25}$$

In particular, note that

$$(w_j - \rho_j) d_j + \gamma_j |d_j - \mu_j| = \begin{cases} (w_j - \rho_j + \gamma_j) d_j - \gamma_j \mu_j, & \text{if } d_j < \mu_j, \\ (w_j - \rho_j) d_j, & \text{if } d_j = \mu_j, \\ (w_j - \rho_j - \gamma_j) d_j + \gamma_j \mu_j, & \text{if } d_j > \mu_j. \end{cases} \tag{C.26}$$

Therefore, (a) if $w_j - \rho_j + \gamma_j \leq 0$, then a maximizer $d_j^\star$ of (C.26) over $\underline{d}_j \leq d_j \leq \overline{d}_j$ is $d_j^\star = \underline{d}_j$; (b) if $w_j - \rho_j - \gamma_j \geq 0$, then a maximizer $d_j^\star$ of (C.26) over $\underline{d}_j \leq d_j \leq \overline{d}_j$ is $d_j^\star = \overline{d}_j$.

We first derive the upper bound on $\rho_{j'}$ for a fixed $j' \in J$. Note that for any $\rho_{j'} - \gamma_{j'} \geq c_{j'}^{\mathrm{u}}$, an optimal solution $\boldsymbol{d}_{j'}^\star$ to the inner supremum problem in (C.25) is $\boldsymbol{d}_{j'}^\star = \underline{d}_{j'}$. Indeed, if $\rho_{j'} - \gamma_{j'} \geq c_{j'}^{\mathrm{u}}$, then it follows from the bound on w_j (i.e., $0 \leq w_j \leq c_j^{\mathrm{u}}$) that $w_{j'} - \rho_{j'} + \gamma_{j'} \leq c_{j'}^{\mathrm{u}} - \rho_{j'} + \gamma_{j'} \leq 0$. Since $\underline{d}_{j'} \leq d_{j'} \leq \overline{d}_{j'}$, an optimal solution $\boldsymbol{d}_{j'}^\star$ to the inner supremum problem in (C.20) is $\boldsymbol{d}_{j'}^\star = \underline{d}_{j'}$. To show the upper bound on $\rho_{j'}$, for any fixed $\boldsymbol{o} \in \mathcal{O}$, let $(\boldsymbol{\rho}^\star,\boldsymbol{\gamma}^\star)$ be an optimal solution to (C.20) with $\rho_{j'}^\star - \gamma_{j'}^\star > c_{j'}^{\mathrm{u}}$. Consider another feasible solution $(\widetilde{\boldsymbol{\rho}},\boldsymbol{\gamma}^\star)$ defined as follows: $\widetilde{\rho}_j = \rho_j^\star$ for $j \in J \setminus \{j'\}$ and $\widetilde{\rho}_{j'} = \rho_j^\star - \varepsilon$ with $\varepsilon \in \left(0, [\rho_{j'}^\star - \gamma_{j'}^\star] - c_{j'}^{\mathrm{u}}\right]$. Note that

$$h(\widetilde{\boldsymbol{\rho}},\boldsymbol{\gamma}^\star;\boldsymbol{o})$$

$$= \boldsymbol{\mu}^\top \widetilde{\boldsymbol{\rho}} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma}^\star + \sup_{(\boldsymbol{w},\boldsymbol{v})\in\Pi,\, \boldsymbol{d}\in\mathcal{S}} \left\{ \sum_{i\in I} C_i o_i v_i + \sum_{j\in J} \left[(w_j - \widetilde{\rho}_j) d_j + \gamma_j^\star |d_j - \mu_j| \right] \right\}$$

$$= \boldsymbol{\mu}^\top \boldsymbol{\rho}^\star - \varepsilon\mu_{j'} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma}^\star + \sup_{(\boldsymbol{w},\boldsymbol{v})\in\Pi,\, \boldsymbol{d}\in\mathcal{S}} \left\{ \sum_{i\in I} C_i o_i v_i + \sum_{j\in J\setminus\{j'\}} \left[(w_j - \rho_j^\star) d_j + \gamma_j^\star |d_j - \mu_j| \right] \right.$$
$$\left. + \left[(w_{j'} - \rho_{j'}^\star + \varepsilon) d_{j'} + \gamma_{j'}^\star |d_{j'} - \mu_{j'}| \right] \right\} \tag{C.27a}$$

$$= \boldsymbol{\mu}^\top \boldsymbol{\rho}^\star - \varepsilon\mu_{j'} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma}^\star + \sup_{(\boldsymbol{w},\boldsymbol{v})\in\Pi,\, \boldsymbol{d}\in\mathcal{S}} \left\{ \sum_{i\in I} C_i o_i v_i + \sum_{j\in J\setminus\{j'\}} \left[(w_j - \rho_j^\star) d_j + \gamma_j^\star |d_j - \mu_j| \right] \right\}$$
$$+ \left[(w_{j'} - \rho_{j'}^\star + \varepsilon) \underline{d}_{j'} + \gamma_{j'}^\star |\underline{d}_{j'} - \mu_{j'}| \right] \tag{C.27b}$$

$$= \boldsymbol{\mu}^\top \boldsymbol{\rho}^\star - \varepsilon\mu_{j'} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma}^\star + \sup_{(\boldsymbol{w},\boldsymbol{v})\in\Pi,\, \boldsymbol{d}\in\mathcal{S}} \left\{ \sum_{i\in I} C_i o_i v_i + \sum_{j\in J} \left[(w_j - \rho_j^\star) d_j + \gamma_j^\star |d_j - \mu_j| \right] \right\} + \varepsilon \underline{d}_{j'} \tag{C.27c}$$

$$= h(\boldsymbol{\rho}^\star, \boldsymbol{\gamma}^\star; \boldsymbol{o}) + \varepsilon(\underline{d}_{j'} - \mu_{j'})$$
$$\leq h(\boldsymbol{\rho}^\star, \boldsymbol{\gamma}^\star; \boldsymbol{o}). \tag{C.27d}$$

Here, (C.27a) follows from the definition of $\widetilde{\boldsymbol{\rho}}$, (C.27b)–(C.27c) follow from the fact that $\underline{d}_{j'}$ is an optimal solution to the supremum problem over $d_{j'}$, (C.27d) follows from $\underline{d}_{j'} \leq \mu_{j'}$. Thus, the new feasible solution $(\widetilde{\boldsymbol{\rho}}, \boldsymbol{\gamma}^\star)$ yields an optimal value no greater than $(\boldsymbol{\rho}^\star, \boldsymbol{\gamma}^\star)$. Thus, without loss of optimality, one can impose the constraint $\rho_{j'} \leq c_{j'}^{\mathrm{u}} + \gamma_{j'}$.

We now derive the lower bound on $\rho_{j'}$ for a fixed $j' \in J$. Note that for any $\rho_{j'} + \gamma_{j'} \leq 0$, an optimal solution $\boldsymbol{d}_{j'}^\star$ to the inner supremum problem in (C.25) is $\boldsymbol{d}_{j'}^\star = \overline{d}_{j'}$. Indeed, if $\rho_{j'} + \gamma_{j'} \leq 0$, then it follows from the bound on w_j (i.e., $0 \leq w_j \leq c_j^{\mathrm{u}}$) that $w_{j'} - \rho_{j'} - \gamma_{j'} \geq -\rho_{j'} - \gamma_{j'} \geq 0$. Since $\underline{d}_{j'} \leq d_{j'} \leq \overline{d}_{j'}$, an optimal solution $\boldsymbol{d}_{j'}^\star$ to the inner supremum problem in (C.20) is $\boldsymbol{d}_{j'}^\star = \overline{d}_{j'}$. To show the lower bound on $\rho_{j'}$, for any fixed $\boldsymbol{o} \in \mathcal{O}$, let $(\boldsymbol{\rho}^\star, \boldsymbol{\gamma}^\star)$ be an optimal solution to (C.20) with $\rho_{j'}^\star + \gamma_{j'}^\star < 0$. Consider another feasible solution $(\widetilde{\boldsymbol{\rho}}, \boldsymbol{\gamma}^\star)$ defined as follows: $\widetilde{\rho}_j = \rho_j^\star$ for $j \in J \setminus \{j'\}$ and $\widetilde{\rho}_{j'} = \rho_j^\star + \varepsilon$ with $\varepsilon \in \left(0, [-\rho_{j'}^\star - \gamma_{j'}^\star]\right]$. Note that

$$h(\widetilde{\boldsymbol{\rho}}, \boldsymbol{\gamma}^\star; \boldsymbol{o})$$
$$= \boldsymbol{\mu}^\top \widetilde{\boldsymbol{\rho}} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma}^\star + \sup_{(\boldsymbol{w},\boldsymbol{v})\in\Pi,\, \boldsymbol{d}\in\mathcal{S}} \left\{ \sum_{i\in I} C_i o_i v_i + \sum_{j\in J} \left[(w_j - \widetilde{\rho}_j) d_j + \gamma_j^\star |d_j - \mu_j| \right] \right\}$$

$$= \boldsymbol{\mu}^\top \boldsymbol{\rho}^\star + \varepsilon\mu_{j'} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma}^\star + \sup_{(\boldsymbol{w},\boldsymbol{v})\in\Pi,\, \boldsymbol{d}\in\mathcal{S}} \left\{ \sum_{i\in I} C_i o_i v_i + \sum_{j\in J\setminus\{j'\}} \left[(w_j - \rho_j^\star) d_j + \gamma_j^\star |d_j - \mu_j| \right] \right.$$
$$\left. + \left[(w_{j'} - \rho_{j'}^\star - \varepsilon) d_{j'} + \gamma_{j'}^\star |d_{j'} - \mu_{j'}| \right] \right\} \tag{C.28a}$$

$$= \boldsymbol{\mu}^\top \boldsymbol{\rho}^\star + \varepsilon\mu_{j'} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma}^\star + \sup_{(\boldsymbol{w},\boldsymbol{v})\in\Pi,\, \boldsymbol{d}\in\mathcal{S}} \left\{ \sum_{i\in I} C_i o_i v_i + \sum_{j\in J\setminus\{j'\}} \left[(w_j - \rho_j^\star) d_j + \gamma_j^\star |d_j - \mu_j| \right] \right\}$$
$$+ \left[(w_{j'} - \rho_{j'}^\star - \varepsilon) \overline{d}_{j'} + \gamma_{j'}^\star |\overline{d}_{j'} - \mu_{j'}| \right] \tag{C.28b}$$

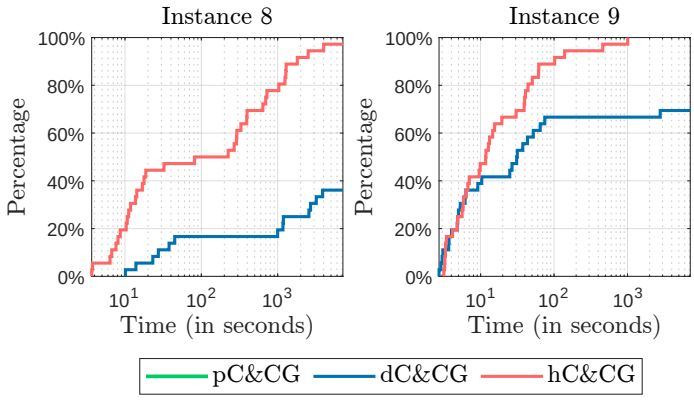


Figure D.12: Time performance profile for pC&CG, dC&CG, and hC&CG when solving instances 8–9 with the MAD ambiguity set and $\varepsilon = 2\%$

$$= \boldsymbol{\mu}^\top \boldsymbol{\rho}^\star + \varepsilon \mu_{j'} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma}^\star + \sup_{(\boldsymbol{w},\boldsymbol{v})\in\Pi,\, \boldsymbol{d}\in\mathcal{S}} \left\{ \sum_{i\in I} C_i o_i v_i + \sum_{j\in J} \left[(w_j - \rho_j^\star) d_j + \gamma_j^\star |d_j - \mu_j| \right] \right\} - \varepsilon \overline{d}_{j'} \tag{C.28c}$$

$$= h(\boldsymbol{\rho}^\star, \boldsymbol{\gamma}^\star; \boldsymbol{o}) + \varepsilon(\mu_{j'} - \overline{d}_{j'})$$

$$\leq h(\boldsymbol{\rho}^\star, \boldsymbol{\gamma}^\star; \boldsymbol{o}). \tag{C.28d}$$

Here, (C.28a) follows from the definition of $\widetilde{\boldsymbol{\rho}}$, (C.28b)–(C.28c) follow from the fact that $\overline{d}_{j'}$ is an optimal solution to the supremum problem over $d_{j'}$, (C.28d) follows from $\overline{d}_{j'} \geq \mu_{j'}$. Thus, the new feasible solution $(\widetilde{\boldsymbol{\rho}}, \boldsymbol{\gamma}^\star)$ yields an optimal value no greater than $(\boldsymbol{\rho}^\star, \boldsymbol{\gamma}^\star)$. Thus, without loss of optimality, one can impose the constraint $\rho_{j'} \geq -\gamma_{j'}$. □

Appendix D. Additional Computational Results

In this section, we provide additional results comparing the computational performance of the proposed hC&CG algorithm with the BD, pC&CG, and dC&CG algorithms for the TRO model with a TRO set constructed using the MAD ambiguity set. We again observe that BD solution times are significantly longer than the C&CG-based algorithms. Thus, we only present results for pC&CG, dC&CG, and hC&CG algorithms. For illustrative purposes, Figure 5 shows the time performance profiles of the three C&CG algorithms for solving instances 8–9. Clearly, hC&CG has the best computational performance among the three algorithms, while pC&CG has the worst. In particular, hC&CG can solve instances 8–9 under nearly all parameter combinations within two hours, while pC&CG cannot solve any. Moreover, hC&CG solution times are considerably shorter than that of dC&CG. These results further demonstrate the superior performance of our new hC&CG algorithm.

References

Ağralı, S., Geunes, J., Taşkın, Z.C., 2012. A facility location model with safety stock costs: Analysis of the cost of single-sourcing requirements. Journal of Global Optimization 54, 551–581.

Ahmadi-Javid, A., Seyedi, P., Syam, S.S., 2017. A survey of healthcare facility location. Computers and Operations Research 79, 223–263.

Albareda-Sambola, M., Fernández, E., Saldanha-da Gama, F., 2017. Heuristic solutions to the facility location problem with general bernoulli demands. INFORMS Journal on Computing 29, 737–753.

Alizadeh, M., Ma, J., Mahdavi-Amiri, N., Marufuzzaman, M., Jaradat, R., 2019. A stochastic programming model for a capacitated location-allocation problem with heterogeneous demands. Computers & Industrial Engineering 137, 106055.

Arslan, A.N., Detienne, B., 2022. Decomposition-based approaches for a class of two-stage robust binary optimization problems. INFORMS journal on computing 34, 857–871.

Ayoub, J., Poss, M., 2016. Decomposition for adjustable robust linear optimization subject to uncertainty polytope. Computational Management Science 13, 219–239.

Bansal, M., Huang, K.L., Mehrotra, S., 2018. Decomposition algorithms for two-stage distributionally robust mixed binary programs. SIAM Journal on Optimization 28, 2360–2383.

Baron, O., Milner, J., Naseraldin, H., 2011. Facility location: A robust optimization approach. Production and Operations Management 20, 772–785.

Basciftci, B., Ahmed, S., Shen, S., 2021. Distributionally robust facility location problem under decision-dependent stochastic demand. European Journal of Operational Research 292, 548–561.

Bertsekas, D., 2009. Convex Optimization Theory. volume 1. Athena Scientific.

Bieniek, M., 2015. A note on the facility location problem with stochastic demands. Omega 55, 53–60.

Blanchet, J., Murthy, K., Nguyen, V.A., 2021. Statistical analysis of Wasserstein distributionally robust estimators, in: Tutorials in Operations Research: Emerging optimization methods and modeling techniques with applications. INFORMS, pp. 227–254.

Blankenship, J.W., Falk, J.E., 1976. Infinitely constrained optimization problems. Journal of Optimization Theory and Applications 19, 261–281.

Cheng, C., Adulyasak, Y., Rousseau, L.M., 2021. Robust facility location under demand uncertainty and facility disruptions. Omega 103, 102429.

Cheng, C., Qi, M., Zhang, Y., Rousseau, L.M., 2018. A two-stage robust approach for the reliable logistics network design problem. Transportation Research Part B: Methodological 111, 185–202.

Cheng, C., Yu, Q., Adulyasak, Y., Rousseau, L.M., 2024. Distributionally robust facility location with uncertain facility capacity and customer demand. Omega 122, 102959.

Conde, E., 2007. Minmax regret location–allocation problem on a network under uncertainty. European Journal of Operational Research 179, 1025–1039.

Daskin, M.S., 2011. Network and Discrete Location: Models, Algorithms, and Applications. John Wiley & Sons.

Daskin, M.S., Coullard, C.R., Shen, Z.J.M., 2002. An inventory-location model: Formulation, solution algorithm and computational results. Annals of Operations Research 110, 83–106.

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

Eiselt, H.A., Marianov, V., 2011. Foundations of Location Analysis. Springer.

Gourtani, A., Nguyen, T.D., Xu, H., 2020. A distributionally robust optimization approach for two-stage facility location problems. EURO Journal on Computational Optimization 8, 141–172.

Güler, O., 2010. Foundations of Optimization. volume 258. Springer Science & Business Media.

Gustafson, S., Kortanek, K., 1973. Numerical treatment of a class of semi-infinite programming problems. Naval Research Logistics Quarterly 20, 477–504.

Jarvis, R.A., 1973. On the identification of the convex hull of a finite set of points in the plane. Information Processing Letters 2, 18–21.

Jiang, R., Zhang, M., Li, G., Guan, Y., 2012. Benders' decomposition for the two-stage security constrained robust unit commitment problem, in: IIE Annual Conference Proceedings, Institute of Industrial and Systems Engineers (IISE). p. 1.

Kayacık, S.E., Basciftci, B., Schrotenboer, A.H., Ursavas, E., 2025. Partially adaptive multistage stochastic programming. European Journal of Operational Research 321, 192–207.

Kuhn, D., Mohajerin Esfahani, P., Nguyen, V.A., Shafieezadeh-Abadeh, S., 2019. Wasserstein distributionally robust optimization: Theory and applications in machine learning, in: Operations Research & Management Science in the Age of Analytics. INFORMS, pp. 130–166.

Laporte, G., Louveaux, F.V., van Hamme, L., 1994. Exact solution to a location problem with stochastic demands. Transportation Science 28, 95–103.

Lei, C., Lin, W.H., Miao, L., 2014. A multicut L-shaped based algorithm to solve a stochastic programming model for the mobile facility routing and scheduling problem. European Journal of Operational Research 238, 699–710.

Lin, C., 2009. Stochastic single-source capacitated facility location model with service level requirements. International Journal of Production Economics 117, 439–451.

Lin, F., Fang, X., Gao, Z., 2022. Distributionally robust optimization: A review on theory and applications. Numerical Algebra, Control and Optimization 12, 159–212.

Liu, K., Li, Q., Zhang, Z.H., 2019. Distributionally robust optimization of an emergency medical service station location and sizing problem with joint chance constraints. Transportation Research Part B: Methodological 119, 79–101.

Liu, T., Saldanha-da Gama, F., Wang, S., Mao, Y., 2022. Robust stochastic facility location: Sensitivity analysis and exact solution. INFORMS Journal on Computing .

Louveaux, F.V., 1986. Discrete stochastic location models. Annals of Operations research 6, 21–34.

Mohajerin Esfahani, P., Kuhn, D., 2018. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. Mathematical Programming 171, 115–166.

Murali, P., Ordonez, F., Dessouky, M.M., 2009. Capacitated facility location with distance dependent coverage under demand uncertainty. Special Issue of Networks and Spatial Economics on Facility Location .

Mutapcic, A., Boyd, S., 2009. Cutting-set methods for robust convex optimization with pessimizing oracles. Optimization Methods & Software 24, 381–406.

Rahimian, H., Mehrotra, S., 2022. Frameworks and results in distributionally robust optimization. Open Journal of Mathematical Optimization 3, 1–85.

Revelle, C.S., Eiselt, H.A., Daskin, M.S., 2008. A bibliography for some fundamental problem categories in discrete location science. European Journal of Operational Research 184, 817–848.

Rockafellar, R.T., 1997. Convex Analysis. volume 11. Princeton University Press.

Saif, A., Delage, E., 2021. Data-driven distributionally robust capacitated facility location problem. European Journal of Operational Research 291, 995–1007.

Santoso, T., Ahmed, S., Goetschalckx, M., Shapiro, A., 2005. A stochastic programming approach for supply chain network design under uncertainty. European Journal of Operational Research 167, 96–115.

Shapiro, A., 2001. On duality theory of conic linear problems. Nonconvex Optimization and its Applications 57, 135–155.

Shehadeh, K.S., Sanci, E., 2021. Distributionally robust facility location with bimodal random demand. Computers & Operations Research 134, 105257.

Shehadeh, K.S., Snyder, L.V., 2023. Equity in stochastic healthcare facility location, in: Uncertainty in Facility Location Problems. Springer, pp. 303–334.

Shehadeh, K.S., Tucker, E.L., 2022. Stochastic optimization models for location and inventory prepositioning of disaster relief supplies. Transportation Research Part C: Emerging Technologies 144, 103871.

Shen, Z.J.M., Coullard, C., Daskin, M.S., 2003. A joint location-inventory model. Transportation Science 37, 40–55.

Shen, Z.J.M., Zhan, R.L., Zhang, J., 2011. The reliable facility location problem: Formulations, heuristics, and approximation algorithms. INFORMS Journal on Computing 23, 470–482.

Sheppard, E.S., 1974. A conceptual framework for dynamic location—allocation analysis. Environment and Planning A 6, 547–564.

Sion, M., 1958. On general minimax theorems. Pacific Journal of Mathematics 8, 171–176.

Smith, J.E., Winkler, R.L., 2006. The optimizer's curse: Skepticism and postdecision surprise in decision analysis. Management Science 52, 311–322.

Subramanyam, A., 2022. A lagrangian dual method for two-stage robust optimization with binary uncertainties. Optimization and Engineering 23, 1831–1871.

Tan, T., Xie, R., Xu, X., Chen, Y., 2024. A robust optimization method for power systems with decision-dependent uncertainty. Energy Conversion and Economics 5, 133–145.

Tanoumand, N., Bodur, M., Naoum-Sawaya, J., 2023. Data-driven distributionally robust optimization: Intersecting ambiguity sets, performance analysis and tractability. Optimization Online https://optimization-online. org .

Thiele, A., Terry, T., Epelman, M., 2009. Robust linear optimization with recourse. Technical Report. Available in Optimization Online.

Tsang, M.Y., Shehadeh, K.S., 2023. Stochastic optimization models for a home service routing and appointment scheduling problem with random travel and service times. European Journal of Operational Research 307, 48–63.

Tsang, M.Y., Shehadeh, K.S., 2025. On the tradeoff between distributional belief and ambiguity: Conservatism, finite-sample guarantees, and asymptotic properties. INFORMS Journal on Optimization .

Tsang, M.Y., Shehadeh, K.S., Curtis, F.E., 2023. An inexact column-and-constraint generation method to solve two-stage robust optimization problems. Operations Research Letters 51, 92–98.

Van Parys, B.P., Mohajerin Esfahani, P., Kuhn, D., 2021. From data to decisions: Distributionally robust optimization is optimal. Management Science 67, 3387–3402.

Wagner, M.R., Bhadury, J., Peng, S., 2009. Risk management in uncapacitated facility location models with random demands. Computers & Operations Research 36, 1002–1011.

Wang, K., Aydemir, M., Jacquillat, A., 2024. Scenario-based robust optimization for two-stage decision making under binary uncertainty. INFORMS Journal on Optimization 6, 84–117.

Wang, K.J., Lee, C.H., 2015. A revised ant algorithm for solving location–allocation problem with risky demand in a multi-echelon supply chain network. Applied Soft Computing 32, 311–321.

Wang, S., Chen, Z., Liu, T., 2020. Distributionally robust hub location. Transportation Science 54, 1189–1210.

Wang, S., Wang, H., Honorio, J., 2023. Learning against distributional uncertainty: On the tradeoff between robustness and specificity. arXiv preprint arXiv:2301.13565 .

Wiesemann, W., Kuhn, D., Sim, M., 2014. Distributionally robust convex optimization. Operations Research 62, 1358–1376.

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

Zeng, B., Zhao, L., 2013. Solving two-stage robust optimization problems using a column-and-constraint generation method. Operations Research Letters 41, 457–461.

Zetina, C.A., Contreras, I., Cordeau, J.F., Nikbakhsh, E., 2017. Robust uncapacitated hub location. Transportation Research Part B: Methodological 106, 393–410.

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

Zhang, Z.H., Berenguer, G., Shen, Z.J., 2015. A capacitated facility location model with bidirectional flows. Transportation Science 49, 114–129.