ALSO-X#: Better Convex Approximations for Distributionally Robust Chance Constrained Programs

Nan Jiang · **Weijun Xie**

January 31, 2023

Abstract This paper studies distributionally robust chance constrained programs (DRCCPs), where the uncertain constraints must be satisfied with at least a probability of a prespecified threshold for all probability distributions from the Wasserstein ambiguity set. As DRCCPs are often nonconvex and challenging to solve optimally, researchers have been developing various convex inner approximations. Recently, $\mathrm{ALSO\text{-}X}$ has been proven to outperform the conditional value-at-risk (CVaR) approximation of a regular chance constrained program when the deterministic set is convex. In this work, we relax this assumption by introducing a new $\mathrm{ALSO\text{-}X\#}$ method for solving DRCCPs. Namely, in the bilevel reformulations of $\mathrm{ALSO\text{-}X}$ and CVaR approximation, we observe that the lower-level $\mathrm{ALSO\text{-}X}$ is a special case of the lower-level CVaR approximation and the upper-level CVaR approximation is more restricted than the one in $\mathrm{ALSO\text{-}X}$. This observation motivates us to propose the $\mathrm{ALSO\text{-}X\#}$, which still resembles a bilevel formulation – in the lower-level problem, we adopt the more general CVaR approximation, and for the upper-level one, we choose the less restricted $\mathrm{ALSO\text{-}X}$. We show that $\mathrm{ALSO\text{-}X\#}$ can always be better than the CVaR approximation and can outperform $\mathrm{ALSO\text{-}X}$ under regular chance constrained programs and type $\infty-$Wasserstein ambiguity set. We also provide new sufficient conditions under which $\mathrm{ALSO\text{-}X\#}$ outputs an optimal solution to a DRCCP. We apply the proposed $\mathrm{ALSO\text{-}X\#}$ to a wireless communication problem and numerically demonstrate that the solution quality can be even better than the exact method.

Keywords. Chance Constraint; CVaR; Distributionally Robust

First Author: Nan Jiang
Affiliation: Georgia Institute of Technology, Atlanta, GA
E-mail: nanjiang@gatech.edu

Corresponding Author: Weijun Xie
Affiliation: Georgia Institute of Technology, Atlanta, GA
E-mail: wxie@gatech.edu

1 Introduction

In this paper, we consider a Distributionally Robust Chance Constrained Program (DRCCP) of form:

$$\text{(DRCCP)}\quad v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} \colon \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} \le b_i(\boldsymbol{x}), \forall i \in [I] \right\} \ge 1-\varepsilon \right\}. \tag{1}$$

In a DRCCP, the objective is to minimize a linear objective function over a deterministic set $\mathcal{X}$ and an uncertain constraint system specified by possibly multiple linear constraints $\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} \le b_i(\boldsymbol{x})$ for all $i \in [I]$, where the uncertain constraints are required to be satisfied with probability $1-\varepsilon$ for any probability distribution $\mathbb{P}$ from an ambiguity set $\mathcal{P}$. Here, the scalar $\varepsilon \in (0,1)$ denotes a preset risk parameter and set $\mathcal{P}$ is formally defined as a subset of probability distributions $\mathbb{P}$ from a measurable space $(\Omega, \mathcal{F})$ equipped with the sigma algebra $\mathcal{F}$ and induced by the random parameters $\tilde{\boldsymbol{\xi}}$ with support set $\Xi \subseteq \mathbb{R}^m$. For each uncertain constraint $i \in [I]$, the affine mappings $\boldsymbol{a}_i(\boldsymbol{x})$ and $b_i(\boldsymbol{x})$ are defined as $\boldsymbol{a}_i(\boldsymbol{x}) = \boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{a}_i \in \mathbb{R}^m$ with $\boldsymbol{A}_i \in \mathbb{R}^{m\times n}, \boldsymbol{a}_i \in \mathbb{R}^m$ and $b_i(\boldsymbol{x}) = \boldsymbol{B}_i^\top \boldsymbol{x} + b_i \in \mathbb{R}$ with $\boldsymbol{B}_i \in \mathbb{R}^n, b_i \in \mathbb{R}$. When there is only $I=1$ uncertain constraint, problem (1) is a *single DRCCP* and otherwise, it is a *joint DRCCP*. Notably, when the ambiguity set $\mathcal{P}$ is a singleton (i.e., $\mathcal{P} = \{\mathbb{P}\}$), DRCCP (1) reduces to a regular Chance Constrained Program (CCP).

1.1 Wasserstein Ambiguity Set

This paper studies the data-driven $q-$Wasserstein ambiguity set defined as

$$\mathcal{P}_q = \left\{ \mathbb{P} \colon \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} \in \Xi \right\} = 1, W_q(\mathbb{P}, \mathbb{P}_{\tilde{\boldsymbol{\zeta}}}) \le \theta \right\},$$

where for any $q \in [1,\infty]$, the $q-$Wasserstein distance is

$$W_q(\mathbb{P}_1,\mathbb{P}_2) = \inf\left\{ \left[\int_{\Xi\times\Xi} \left\|\boldsymbol{\xi}^1 - \boldsymbol{\xi}^2\right\|^q \mathbb{Q}(d\boldsymbol{\xi}^1, d\boldsymbol{\xi}^2) \right]^{\frac{1}{q}} \colon \begin{array}{c} \mathbb{Q} \text{ is a joint distribution of } \tilde{\boldsymbol{\xi}}^1 \text{ and } \tilde{\boldsymbol{\xi}}^2 \\ \text{with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2, \text{ respectively} \end{array} \right\},$$

$\theta \ge 0$ is the Wasserstein radius, and $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ denotes the reference distribution induced by random parameters $\tilde{\boldsymbol{\zeta}}$. Recently, there are many exciting works on DRCCP under type $q-$Wasserstein ambiguity set [11, 12, 22, 25, 26, 29, 45, 49]. Particularly, according to the equivalent reformulation in proposition 8 of [26], we write DRCCP (1) under type $\infty-$Wasserstein ambiguity set as

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \mathbb{P}\left\{ \tilde{\boldsymbol{\zeta}} \colon \theta\left\|\boldsymbol{a}_i(\boldsymbol{x})\right\|_* + \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} \le b_i(\boldsymbol{x}), \forall i \in [I] \right\} \ge 1-\varepsilon \right\}. \tag{2}$$

Throughout the paper, we make the following assumption:

A1 The reference distribution $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ is sub-Gaussian, that is, $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}\{\tilde{\boldsymbol{\zeta}} : \|\tilde{\boldsymbol{\zeta}}\| \ge \tau\} \le C_1 \exp(-C_2\tau^2)$ for some positive constants C_1, C_2.

It is worth noting that the sub-Gaussian assumption ensures the weak compactness of the Wasserstein ambiguity set and thus enjoys the strong duality of reformulating the worst-case expectation under type $q-$Wasserstein ambiguity set. Particularly, this paper mainly focuses on empirical or elliptical reference distributions, which clearly satisfy Assumption A1.

1.2 Relevant Literature

Distributionally robust chance constrained programs (DRCCPs) have gained much attention recently when the knowledge about the probability distribution is limited (see, e.g., [11, 12, 20–23, 25, 26, 29, 44, 45, 49, 50, 54]). As DRCCPs' feasible regions are often nonconvex, some existing research has worked on identifying conditions under which the feasible region in DRCCP (1) is convex (see, e.g., [10, 14, 16, 20, 30, 32, 40, 43, 44, 50]). For a single DRCCP (1), the authors in [44] showed that its feasible region is convex if the reference distribution is Gaussian under type $1-$Wasserstein ambiguity set. Similar convexity results apply to a single DRCCP when the ambiguity set $\mathcal{P}$ comprises all probability distributions with known first and second moments (see, e.g., [10, 16]), known support of $\tilde{\boldsymbol{\xi}}$ (see, e.g., [14]), arbitrary convex mapping of $\tilde{\boldsymbol{\xi}}$ (see, e.g., [50]), or the unimodality property of $\mathbb{P}$ (see, e.g., [20, 32]). Researchers have also proposed convex inner approximations of the nonconvex chance constraint (e.g., [1, 9, 12, 26, 35, 36]). For example, the well-known conditional value-at-risk (CVaR) approximation is to replace the chance constraint in DRCCP (1) with the more conservative CVaR constraint (see the details in [35]). Recently, ALSO-X, a convex approximation method proposed in [1], has been proven to outperform the CVaR approximation of a regular chance constrained program (see, e.g., theorem 1 in [26]).

Despite the challenges, DRCCPs are effective in decision-making under uncertainty and have been applied to a wide range of problems, including portfolio optimization [37, 38], energy management [8, 46], supply chain management [17, 18], facility location problems [31], and wireless communication network [4, 33, 34, 48]. For example, chance constraints have been used in the design and optimization of wireless communication networks to ensure that the probability of certain operational constraints being violated, such as capacity limits or reliability requirements, is within an acceptable limit (see, e.g., [33, 34]). In portfolio optimization, the objective is to maximize the expected return of the portfolio while ensuring that the probability of portfolio losses does not exceed a specified level (see, e.g., [37, 38]). We refer interested readers to [2] for more applications. For a comprehensive review of DRCCPs, interested readers are referred to a recent survey from [29].

1.3 Contributions

In this paper, we study a new method, termed "ALSO-X#," which advances the recent ALSO-X in [26] in the following three main aspects: (a) for any closed deterministic set $\mathcal{X}$, ALSO-X# is always better than CVaR approximation under any ambiguity set; (b) ALSO-X# can be better than ALSO-X; and (c) ALSO-X# admits new conditions under which its output solution is also optimal to DRCCPs. More specifically,

(i) We prove that ALSO-X# is better than CVaR approximation under a general ambiguity set (beyond Wasserstein ambiguity set) and a closed deterministic set $\mathcal{X}$. This result significantly improves that of theorem 1 in [26]), which relies on the convexity of the deterministic set $\mathcal{X}$;
(ii) We show that under type $\infty-$ Wasserstein ambiguity set, ALSO-X# is better than ALSO-X when the lower-level ALSO-X admits a unique solution. When the reference distribution is constructed by i.i.d. samples from a continuous nondegerate distribution, or the reference distribution is continuous and nondedegerate, the lower-level ALSO-X indeed presents a unique solution;
(iii) We present new sufficient conditions under which ALSO-X# yields an optimal solution to a DR-CCP. For example, one sufficient condition is that for a binary DRCCP with an empirical reference distribution; and
(iv) We extend the afromentioned results of ALSO-X# to solve a DRCCP (1) under type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$.

Organization. The remainder of the paper is organized as follows. Section 2 reviews ALSO-X and CVaR approximation and introduces the ALSO-X#. Section 3 shows that ALSO-X# is better than ALSO-X and CVaR approximation. Section 4 explores conditions under which ALSO-X# is better than ALSO-X.

Section 5 provides the conditions under which ALSO-X# outputs an exact optimal solution. Section 6 extends ALSO-X# to solve DRCCPs under type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$. Section 7 numerically illustrates the proposed methods. Section 8 concludes the paper.

Notation. The following notation is used throughout the paper. We use bold letters (e.g., $\boldsymbol{x}, \boldsymbol{A}$) to denote vectors and matrices and use corresponding non-bold letters to denote their components. Given a vector or matrix $\boldsymbol{x}$, its zero norm $\|\boldsymbol{x}\|_0$ denotes the number of its nonzero elements. We let $\|\cdot\|_*$ denote the dual norm of a general norm $\|\cdot\|$. Given an integer n, we let $[n] := \{1, 2, \cdots, n\}$ and use $\mathbb{R}_+^n := \{\boldsymbol{x} \in \mathbb{R}^n : x_i \geq 0, \forall i \in [n]\}$. Given a real number τ, let $(\tau)_+ := \max\{\tau, 0\}$. Given a finite set I, let $|I|$ denote its cardinality. We let $\tilde{\boldsymbol{\xi}}$ denote a random vector and denote its realizations by $\boldsymbol{\xi}$. Given a vector $\boldsymbol{x} \in \mathbb{R}^n$, let $\operatorname{supp}(\boldsymbol{x})$ be its support, i.e., $\operatorname{supp}(\boldsymbol{x}) := \{i \in [n] : x_i \neq 0\}$. Given a probability distribution $\mathbb{P}$ on Ξ, we use $\mathbb{P}\{A\}$ to denote $\mathbb{P}\{\boldsymbol{\xi} : \text{condition } A(\boldsymbol{\xi}) \text{ holds}\}$ when $A(\boldsymbol{\xi})$ is a condition on $\boldsymbol{\xi}$, and to denote $\mathbb{P}\{\boldsymbol{\xi} \colon \boldsymbol{\xi} \in A\}$ when $A \subseteq \Xi$ is $\mathbb{P}-$measurable. We use $\lfloor x \rfloor$ to denote the largest integer y satisfying $y \leq x$, for any $x \in \mathbb{R}$. We use the phrase "Better Than" to indicate "at least as good as." Additional notations will be introduced as needed.

2 ALSO-X#

In this section, we first review two convex approximations of DRCCP, ALSO-X and CVaR approximation. Then we present ALSO-X# for solving DRCCP (2) and show its connections to ALSO-X and CVaR approximation. To begin with, we first introduce the notions of $\mathrm{VaR}_{1-\varepsilon}(\cdot)$ and $\mathrm{CVaR}_{1-\varepsilon}(\cdot)$. For a given risk parameter ε and a given random variable $\tilde{\boldsymbol{X}}$ with probability distribution $\mathbb{P}$ and cumulative distribution function $F_{\tilde{\boldsymbol{X}}}(\cdot)$, $(1-\varepsilon)$ Value-at-Risk (VaR) of $\tilde{\boldsymbol{X}}$ is defined as

$$\mathrm{VaR}_{1-\varepsilon}(\tilde{\boldsymbol{X}}) := \min_{s} \left\{ s : F_{\tilde{\boldsymbol{X}}}(s) \geq 1-\varepsilon \right\},$$

and the corresponding Conditional Value-at-Risk (CVaR) is

$$\mathrm{CVaR}_{1-\varepsilon}(\tilde{\boldsymbol{X}}) := \min_{\beta} \left\{ \beta + \frac{1}{\varepsilon} \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{X}} - \beta]_+ \right\}.$$

2.1 State-of-the-art Convex Approximations

In general, solving DRCCP (2) is NP-hard (see, e.g., [51]). Thus, in this work, instead of solving DRCCP (2) directly, we review two known convex approximations, i.e., the popular CVaR approximation and the recent ALSO-X.

The ALSO-X method with a bilevel structure can be adapted to solve DRCCP (2). In the lower-level ALSO-X, we solve the hinge-loss approximation with a given objective upper bound. We then check whether its optimal solution $\boldsymbol{x}^*$ satisfies the worst-case chance constraint or not. The upper-level ALSO-X is to search the best upper bound of the objective value. Formally, ALSO-X admits the form:

$$\begin{aligned} v^A = \min_{t} \quad & t, \\ \text{s.t.} \quad & \boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{P}_\infty} \left\{ \mathbb{E}_{\mathbb{P}} \left[\max_{i \in [I]} \left(\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} - b_i(\boldsymbol{x}) \right)_+ \right] : \boldsymbol{c}^\top \boldsymbol{x} \leq t \right\}, \\ & \inf_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P} \left\{ \tilde{\boldsymbol{\xi}} \colon \boldsymbol{a}_i(\boldsymbol{x}^*)^\top \tilde{\boldsymbol{\xi}} \leq b_i(\boldsymbol{x}^*), \forall i \in [I] \right\} \geq 1-\varepsilon. \end{aligned}$$

Based on the equivalent reformulation in proposition 9 of [26], we consider the following ALSO-X:

$$v^A = \min_{t} \quad t, \tag{3a}$$

$$\text{s.t.} \quad \boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} \left\{ \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left[\max_{i\in[I]} \left(\theta \left\| \boldsymbol{a}_i(\boldsymbol{x}) \right\|_* + \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) \right)_+ \right] : \boldsymbol{c}^\top \boldsymbol{x} \le t \right\}, \tag{3b}$$

$$\mathbb{P}\left\{ \tilde{\boldsymbol{\zeta}} \colon \theta \left\| \boldsymbol{a}_i(\boldsymbol{x}^*) \right\|_* + \boldsymbol{a}_i(\boldsymbol{x}^*)^\top \tilde{\boldsymbol{\zeta}} \le b_i(\boldsymbol{x}^*), \forall i \in [I] \right\} \ge 1 - \varepsilon. \tag{3c}$$

Under type $\infty-$Wasserstein ambiguity set $\mathcal{P}_\infty$, it has been shown that when the deterministic set $\mathcal{X}$ is convex, ALSO-X is better than CVaR approximation (see, e.g., theorem 7 in [26]). However, this result, in general, does not hold for a DRCCP when set $\mathcal{X}$ is nonconvex (see example 2 in [26] with Wasserstein radius $\theta = 0$). For notational convenience, let us denote $v^A(t)$ and $\widehat{F}(\boldsymbol{x})$ to be the optimal value and the objective function of the lower-level ALSO-X (3b), respectively.

The CVaR approximation has been shown to work quite well for solving DRCCPs (see, e.g., [11, 49]). For DRCCP (2), its CVaR approximation is defined by replacing the worst-case chance constraint by the worst-case CVaR constraint as below

$$v^{\mathrm{CVaR}} = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \sup_{\mathbb{P}\in\mathcal{P}_\infty} \inf_{\beta\le 0} \left[\beta + \frac{1}{\varepsilon} \mathbb{E}_{\mathbb{P}} \left[\left(\max_{i\in[I]} \left(\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} - b_i(\boldsymbol{x}) \right) - \beta \right)_+ \right] \right] \le 0 \right\}.$$

From the equivalent reformulation in proposition 9 of [26], we consider the following CVaR approximation of DRCCP (2):

$$v^{\mathrm{CVaR}} = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \inf_{\beta\le 0} \left[\beta + \frac{1}{\varepsilon} \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left[\left(\max_{i\in[I]} \left(\theta \left\| \boldsymbol{a}_i(\boldsymbol{x}) \right\|_* + \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) \right) - \beta \right)_+ \right] \right] \le 0 \right\}.$$

Equivalently, we also recast the CVaR approximation as a bilevel program, where the upper-level problem is to search the best objective value and the lower-level problem is to minimize the left-hand of CVaR constraint given that the objective function is upper-bounded by a given value. That is,

$$v^{\mathrm{CVaR}} = \min_t t, \tag{4a}$$

$$\text{s.t.} \ (\boldsymbol{x}^*, \beta^*) \in \operatorname*{argmin}_{\substack{\boldsymbol{x}\in\mathcal{X}, \boldsymbol{c}^\top \boldsymbol{x} \le t, \\ \beta \le 0}} \left\{ \varepsilon\beta + \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left[\left(\max_{i\in[I]} \left(\theta \left\| \boldsymbol{a}_i(\boldsymbol{x}) \right\|_* + \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) \right) - \beta \right)_+ \right] \right\}, \tag{4b}$$

$$\varepsilon\beta^* + \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left[\left(\max_{i\in[I]} \left(\theta \left\| \boldsymbol{a}_i(\boldsymbol{x}^*) \right\|_* + \boldsymbol{a}_i(\boldsymbol{x}^*)^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}^*) \right) - \beta^* \right)_+ \right] \le 0. \tag{4c}$$

We call the objective function in the lower-level CVaR approximation (4b) "CVaR-loss." We observe that if we let variable $\beta = 0$ in the lower-level CVaR approximation (4b), then we recover the hinge-loss approximation (3b). In other words, the lower-level ALSO-X (3b) and the lower-level CVaR approximation (4b) coincide when $\beta = 0$. This observation inspires us to improve ALSO-X by replacing its lower-level hinge-loss objective function with the CVaR-loss approximation, which is elaborated in the subsequent subsections.

It is worth noting that, when the deterministic set $\mathcal{X}$ is discrete, CVaR approximation can outperform ALSO-X when solving DRCCP (2), as demonstrated by the following example.

Example 1 Consider a single DRCCP under type $\infty-$Wasserstein ambiguity set with $\theta = 1$. Assume that the empirical distribution has 4 equiprobable scenarios (i.e., $N = 4$, $\mathbb{P}\{\tilde{\boldsymbol{\zeta}} = \boldsymbol{\zeta}^i\} = 1/N$), risk parameter $\varepsilon = 1/2$, deterministic set $\mathcal{X} = \{0, 1\}$, function $\boldsymbol{a}_1(x)^\top \boldsymbol{\zeta} - b_1(x) = \zeta_1 x - \zeta_2$, $\zeta_1^1 = -48$, $\zeta_1^2 = \zeta_1^3 = \zeta_1^4 = 100$, $\zeta_2^1 = -51$, and $\zeta_2^2 = \zeta_2^3 = \zeta_2^4 = 100$. In this example, DRCCP (2) resorts to

$$v^* = \min_{x\in\{0,1\}} \left\{ -x \colon \mathbb{I}(49x \ge 50) + \mathbb{I}(101x \le 99) + \mathbb{I}(101x \le 99) + \mathbb{I}(101x \le 99) \ge 2 \right\},$$

where CVaR approximation returns the optimal solution and ALSO-X fails to find any feasible solution (see, example 2 in [26]). $\diamond$

2.2 What is ALSO-X#?

To overcome the limitations of ALSO-X and CVaR approximation, we introduce the new "ALSO-X#." As discussed in the previous subsection, the lower-level ALSO-X (3b) can be viewed as a special case of the lower-level CVaR approximation (4b) by letting $\beta = 0$. Thus, one may want to replace the hinge-loss objective function with the CVaR-loss one. On the other hand, one disadvantage of the CVaR approximation (4) is that it relies on a more conservative CVaR constraint (4c) for the feasibility check. To improve the CVaR approximation (4), we can use chance constraint (3c) for the feasibility check, leading to ALSO-X#, an integration of CVaR approximation and ALSO-X. Formally, ALSO-X# admits the form of

$$
\begin{aligned}
v^{A\#} = \min_{t} \quad & t, \\
\text{s.t.} \quad & (\boldsymbol{x}^*, \beta^*) \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}, \beta\le 0} \sup_{\mathbb{P}\in\mathcal{P}_\infty} \left\{ \varepsilon\beta + \mathbb{E}_{\mathbb{P}}\left[\max_{i\in[I]} \left(\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} - b_i(\boldsymbol{x}) \right) - \beta \right]_+ : \boldsymbol{c}^\top \boldsymbol{x} \le t \right\}, \\
& \inf_{\mathbb{P}\in\mathcal{P}_\infty} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} \colon \boldsymbol{a}_i(\boldsymbol{x}^*)^\top \tilde{\boldsymbol{\xi}} \le b_i(\boldsymbol{x}^*), \forall i \in [I] \right\} \ge 1 - \varepsilon.
\end{aligned}
$$

According to the reformulations in Section 2.1, the ALSO-X# is equivalent to

$$
v^{A\#} = \min_{t} \quad t, \tag{5a}
$$

$$
\text{s.t.} \quad (\boldsymbol{x}^*, \beta^*) \in \operatorname*{argmin}_{\substack{\boldsymbol{x}\in\mathcal{X}, \boldsymbol{c}^\top\boldsymbol{x}\le t,\\ \beta\le 0}} \left\{ \varepsilon\beta + \mathbb{E}_{\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}}\left[\left(\max_{i\in[I]} \left(\theta \left\| \boldsymbol{a}_i(\boldsymbol{x}) \right\|_* + \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) \right) - \beta \right)_+ \right] \right\}, \tag{5b}
$$

$$
\mathbb{P}\left\{ \tilde{\boldsymbol{\zeta}} \colon \theta \left\| \boldsymbol{a}_i(\boldsymbol{x}^*) \right\|_* + \boldsymbol{a}_i(\boldsymbol{x}^*)^\top \tilde{\boldsymbol{\zeta}} \le b_i(\boldsymbol{x}^*), \forall i \in [I] \right\} \ge 1 - \varepsilon. \tag{5c}
$$

In the proposed ALSO-X# (5), given a current objective value t, the lower-level ALSO-X# (5b) is to solve the CVaR approximation first and then the upper-level ALSO-X# is to check whether the lower-level solution satisfies the worst-case chance constraint (5c) or not. In this way, we introduce the new convex approximation of DRCCP (2), where we have $v^* \le v^{A\#}$.

Algorithm 1 summarizes the solution procedure of ALSO-X# (5), where for a given t of the upper-level problem, we solve the lower-level CVaR approximation (5b) with an optimal solution $(\boldsymbol{x}^*, \beta^*)$ and check whether $\boldsymbol{x}^*$ is feasible to DRCCP (2) or not, i.e., check if $\boldsymbol{x}^*$ satisfies (5c) or not. If the answer is YES, we decrease the value of t; otherwise, increase it. In the implementation, we search the optimal t by using the binary search method with a stopping tolerance δ_1. The implementation details follow similarly to algorithm 1 in [26] and the remarks therein.

Algorithm 1 The Proposed ALSO-X# Algorithm

1: **Input:** Let δ_1 denote the stopping tolerance parameter, t_L and t_U be the known lower and upper bounds of the optimal value of DRCCP (2), respectively
2: **while** $t_U - t_L > \delta_1$ **do**
3: Let $t = (t_L + t_U)/2$ and $(\boldsymbol{x}^*, \beta^*)$ be an optimal solution of the lower-level ALSO-X# (5b)
4: Let $t_L = t$ if $\boldsymbol{x}^*$ satisfies (5c); otherwise, $t_U = t$
5: **end while**
6: **Output:** A feasible solution $\boldsymbol{x}^*$ and its objective value $\bar{v}^{A\#}$ to DRCCP (2)

Note that we inherit the constraint $\beta \le 0$ in the lower-level ALSO-X# (5b) from the CVaR approximation. One may want to relax this constraint and arrive at the following "weak" formulation of

ALSO-X#, termed as "ALSO-X$\underline{\#}$:"

$$v^{A\underline{\#}} = \min_{t} t, \tag{6a}$$

$$\text{s.t. } (\boldsymbol{x}^*, \beta^*) \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}, \boldsymbol{c}^\top \boldsymbol{x} \le t, \beta} \left\{ \varepsilon\beta + \mathbb{E}_{\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}} \left[\left(\max_{i\in[I]} \left(\theta \left\| \boldsymbol{a}_i(\boldsymbol{x}) \right\|_* + \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) \right) - \beta \right)_+ \right] \right\}, \tag{6b}$$

$$\mathbb{P}\left\{ \tilde{\boldsymbol{\zeta}} \colon \theta \left\| \boldsymbol{a}_i(\boldsymbol{x}^*) \right\|_* + \boldsymbol{a}_i(\boldsymbol{x}^*)^\top \tilde{\boldsymbol{\zeta}} \le b_i(\boldsymbol{x}^*), \forall i \in [I] \right\} \ge 1 - \varepsilon. \tag{6c}$$

In our numerical study, we find that ALSO-X# consistently outperforms ALSO-X$\underline{\#}$. The following example shows that ALSO-X# can be superior to ALSO-X$\underline{\#}$, ALSO-X, ALSO-X+, and CVaR approximation. We formally prove this result in the next section.

Example 2 Consider a single DRCCP under type $\infty-$Wasserstein ambiguity set with $\theta = 1$. Assume that the empirical distribution has 4 equiprobable scenarios (i.e., $N = 4$, $\mathbb{P}\{\tilde{\boldsymbol{\zeta}} = \boldsymbol{\zeta}^i\} = 1/N$), risk parameter $\varepsilon = 1/2$, deterministic set $\mathcal{X} = \{0, 1\}$, function $\boldsymbol{a}_1(x)^\top \boldsymbol{\zeta} - b_1(x) = \zeta_1 x - \zeta_2$, $\zeta_1^1 = -8$, $\zeta_1^2 = \zeta_1^3 = \zeta_1^4 = 3$, $\zeta_2^1 = -25/2$, and $\zeta_2^2 = \zeta_2^3 = \zeta_2^4 = 5/2$. In this example, DRCCP (2) resorts to

$$v^* = \min_{x\in\{0,1\}} \left\{ -x \colon \mathbb{I}\left(9x \ge \frac{23}{2}\right) + \mathbb{I}\left(4x \le \frac{3}{2}\right) + \mathbb{I}\left(4x \le \frac{3}{2}\right) + \mathbb{I}\left(4x \le \frac{3}{2}\right) \ge 2 \right\}.$$

The weak formulation of ALSO-X# $v^{A\underline{\#}}$ (6) can be written as

$$v^{A\underline{\#}} = \min_{t} \left\{ t \colon \sum_{i\in[4]} \mathbb{I}(s_i^* > 0) \le 2, \right.$$

$$\left. (x^*, \boldsymbol{s}^*, \beta^*) \in \operatorname*{argmin}_{x\in\{0,1\}, \boldsymbol{s}, \beta} \left\{ \frac{1}{4} \sum_{i\in[4]} s_i - \frac{1}{2}\beta \colon \begin{array}{l} -9x + \dfrac{23}{2} \le s_1, 4x - \dfrac{3}{2} \le s_2, 4x - \dfrac{3}{2} \le s_3, \\ 4x - \dfrac{3}{2} \le s_4, -x \le t, s_i \ge \beta, \forall i \in [4] \end{array} \right\} \right\}.$$

Particularly, for any $t \ge -1$, the ALSO-X$\underline{\#}$ returns a solution with $s_1^* = s_2^* = s_3^* = s_4^* = 5/2$, $x^* = 1$. Since the support size of $\boldsymbol{s}^*$ is greater than 2, we have to increase the objective bound t to be infinite. However, if we enforce $\beta \le 0$, i.e., consider the ALSO-X#, we have $s_1^* = 23/2, s_2^* = s_3^* = s_4^* = -3/2, \beta^* = -3/2, x^* = 0$. Therefore, in this example, ALSO-X# always returns the optimal solution, but ALSO-X$\underline{\#}$ fails to find any feasible solution. Notice that in this example, ALSO-X, CVaR approximation, and ALSO-X+ (see, e.g., section 4 in [26]) all fail to find any feasible solution. $\diamond$

It is worthy of mentioning that albeit the formulation of ALSO-X$\underline{\#}$ tends to be weaker, the unconstrained β variable is useful to prove the strength of ALSO-X# under an elliptical reference distribution.

3 ALSO-X# is Better Than ALSO-X and CVaR Approximation

In this section, we prove that ALSO-X# is better than ALSO-X, ALSO-X$\underline{\#}$, and CVaR approximation, respectively.

3.1 ALSO-X# is Better Than ALSO-X

Note that ALSO-X# can be viewed as an integration of CVaR approximation and ALSO-X. In fact, if the lower-level ALSO-X provides a feasible solution to DRCCP (2), then at least one optimal solution of the lower-level ALSO-X# is also feasible to DRCCP (2). Particularly, under type $\infty-$Wasserstein ambiguity set and for a given objective upper bound t such that the optimal value of the lower-level ALSO-X (3b) is positive (i.e., $v^A(t) > 0$), the lower-level ALSO-X (3b) has a unique optimal solution, which is feasible to DRCCP (2). Then any optimal solution of the lower-level ALSO-X# is also feasible to DRCCP (2).

Theorem 1 *Suppose that for any objective upper bound t such that $t \geq \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and $v^A(t) > 0$, the lower-level* ALSO-X (3b) *admits a unique optimal solution. Then* ALSO-X# (5) *is better than* ALSO-X (3), *i.e.,* $v^{A\#} \leq v^A$.

Proof It is sufficient to show that for any objective upper bound t, if the lower-level ALSO-X (3b) is feasible to DRCCP (2), then the lower-level ALSO-X# (5b) is also feasible to DRCCP (2). Let $(\widehat{\boldsymbol{x}}, \widehat{\beta})$ denote an optimal solution from the lower-level ALSO-X# (5b) and let $\bar{\boldsymbol{x}}$ denote an optimal solution from the lower-level ALSO-X (3b). We split the proof into two steps by discussing $v^A(t) = 0$ or $v^A(t) > 0$.

Step I. When the optimal value of the lower-level ALSO-X (3b) is $v^A(t) = 0$, since $v^A(t)$ is an upper bound of the lower-level ALSO-X# (5b). Thus, the optimal value of the lower-level ALSO-X (3b) is less than or equal to zero. Due to the fact that the objective value of ALSO-X# divided by ε less than or equal to zero implies a conservative approximation of distributionally robust chance constraint [36], we must have that $\widehat{\boldsymbol{x}}$ is feasible to DRCCP (2).

Step II. Next, we consider the case when the optimal value from the lower-level ALSO-X (3b) is positive, i.e., $v^A(t) > 0$. We discuss two cases on whether $\widehat{\beta} = 0$ or not.

Case I. If $\widehat{\beta} = 0$, then the lower-level ALSO-X# (5b) and the lower-level ALSO-X (3b) coincide. Since the lower-level ALSO-X (3b) admits a unique optimal solution, we must have $\bar{\boldsymbol{x}} = \widehat{\boldsymbol{x}}$. That is, ALSO-X# and ALSO-X are the same. Thus, if $\bar{\boldsymbol{x}}$ is feasible to the DRCCP, i.e., $\bar{\boldsymbol{x}}$ satisfies (3c), $\widehat{\boldsymbol{x}}$ is also feasible to DRCCP (2).

Case II. Suppose that $\widehat{\beta} < 0$. From the discussions in section 3 of [41], we have

$$\mathrm{VaR}_{1-\varepsilon}\left\{\theta \left\|\boldsymbol{a}_i(\widehat{\boldsymbol{x}})\right\|_* + \boldsymbol{a}_i(\widehat{\boldsymbol{x}})^\top \tilde{\boldsymbol{\zeta}} - b_i(\widehat{\boldsymbol{x}}), \forall i \in [I]\right\} < 0,$$

which implies that

$$\mathbb{P}\left\{\tilde{\boldsymbol{\zeta}} \colon \theta \left\|\boldsymbol{a}_i(\widehat{\boldsymbol{x}})\right\|_* + \boldsymbol{a}_i(\widehat{\boldsymbol{x}})^\top \tilde{\boldsymbol{\zeta}} \leq b_i(\widehat{\boldsymbol{x}}), \forall i \in [I]\right\} \geq 1 - \varepsilon.$$

Thus, ALSO-X# provides a feasible solution to DRCCP (2). Since the solution of the lower-level ALSO-X (3b) may not be feasible to DRCCP (2), ALSO-X# is better than ALSO-X. ☐

We make the following remarks about Theorem 1:

(i) Different from the work [26], our proof does not require the convexity assumption of set $\mathcal{X}$;
(ii) The uniqueness condition is satisfied by many DRCCPs as well as their regular counterparts, as formally proved in the next section; and
(iii) In Example 2, the lower-level ALSO-X has a unique optimal solution when $t \geq -1$, which, however, is not feasible to the DRCCP. On the contrary, the proposed ALSO-X# can find the optimal solution, which is better than ALSO-X according to Theorem 1.

The uniqueness assumption in Theorem 1 is, in fact, necessary. Below is an example showing that ALSO-X# can be worse than ALSO-X.

Example 3 Consider a single DRCCP under type $\infty-$Wasserstein ambiguity set with $\theta = 1/2$ and $\|\cdot\|_* = \|\cdot\|_1$. Assume that the empirical distribution has 3 equiprobable scenarios (i.e., $N = 3$, $\mathbb{P}\{\tilde{\boldsymbol{\zeta}} = \boldsymbol{\zeta}^i\} = 1/N$), risk parameter $\varepsilon = 1/2$, deterministic set $\mathcal{X} = [0, 10]^2$, function $\boldsymbol{a}_1(\boldsymbol{x})^\top \boldsymbol{\zeta} - b_1(\boldsymbol{x}) = -\boldsymbol{x}^\top \boldsymbol{\zeta} + 1$, $\boldsymbol{\zeta}^1 = (5/2, 7/2)^\top$, $\boldsymbol{\zeta}^2 = (5/2, 3/2)^\top$, and $\boldsymbol{\zeta}^3 = (3/2, 5/2)^\top$. In this example, DRCCP (2) resorts to

$$v^* = \min_{\boldsymbol{x}\in[0,10]^2} \left\{x_1 + x_2 \colon \mathbb{I}(2x_1 + 3x_2 \geq 1) + \mathbb{I}(2x_1 + x_2 \geq 1) + \mathbb{I}(x_1 + 2x_2 \geq 1) \geq 2\right\},$$

where the optimal value $v^* = 1/2$. Its ALSO-X counterpart admits the following form:

$$v^A = \min_t \Bigg\{t \colon \sum_{i\in[3]} \mathbb{I}(s_i^* > 0) \leq 2,$$

$$(\boldsymbol{x}^*, \boldsymbol{s}^*) \in \underset{\boldsymbol{x}\in[0,10]^2, \boldsymbol{s}\in\mathbb{R}^3_+}{\operatorname{argmin}} \left\{ \frac{1}{3}\sum_{i\in[3]} s_i : \begin{array}{l} 2x_1+3x_2 \geq 1-s_1, 2x_1+x_2\geq 1-s_2, \\ x_1+2x_2\geq 1-s_3, x_1+x_2 \leq t \end{array} \right\} \Bigg\}.$$

When $t = 1/2$, one optimal solution of the lower-level ALSO-X is $x_1^* = 0, x_2^* = 1/2, s_1^* = 0, s_2^* = 1/2, s_3^* = 0$, which is feasible to the DRCCP. In this case, ALSO-X can find an optimal solution of the DRCCP with $v^A = v^* = 1/2$. However, when $t = 1/2$, another optimal solution of the lower-level ALSO-X is $x_1^* = 1/4, x_2^* = 1/4, s_1^* = 0, s_2^* = 1/4, s_3^* = 1/4$, which is infeasible to the DRCCP. Hence, it violates the uniqueness assumption.

Now let us consider corresponding ALSO-X#:

$$v^{A\#} = \min_t \Bigg\{ t\colon \sum_{i\in[3]} \mathbb{I}(s_i^* > 0) \leq 2,$$

$$(\boldsymbol{x}^*, \boldsymbol{s}^*, \beta^*) \in \underset{\boldsymbol{x}\in[0,10]^2, \boldsymbol{s}, \beta\leq 0}{\operatorname{argmin}} \left\{ \frac{1}{3}\sum_{i\in[3]} s_i - \frac{1}{2}\beta : \begin{array}{l} 2x_1+3x_2 \geq 1-s_1, 2x_1+x_2\geq 1-s_2, \\ x_1+2x_2\geq 1-s_3, x_1+x_2 \leq t, s_i\geq\beta, \forall i\in[3] \end{array} \right\} \Bigg\}.$$

When $t = 1/2$, one of its optimal solution is $x_1^* = 1/4, x_2^* = 1/4, s_1^* = 0, s_2^* = 1/4, s_3^* = 1/4, \beta^* = 0$, which is infeasible to the DRCCP. Thus, ALSO-X# may not be able to find an optimal solution of the DRCCP. That is, we can have $v^{A\#} > v^A$ when the uniqueness assumption is violated. ◇

3.2 ALSO-X# is Better Than ALSO-X$\underline{\#}$

In the lower-level ALSO-X# (5b), we impose the constraint $\beta \leq 0$ based on the CVaR approximation (4). By relaxing this constraint, we obtain a weaker ALSO-X$\underline{\#}$ (6). We show that when ALSO-X$\underline{\#}$ provides a feasible solution to DRCCP (2), the lower-level ALSO-X# is equivalent to that of ALSO-X$\underline{\#}$. Particularly, under type $\infty-$Wasserstein ambiguity set and for a given objective upper bound t, when there exists an optimal solution from the lower-level ALSO-X$\underline{\#}$ (6b) that is feasible to DRCCP (2), any optimal solution of the lower-level ALSO-X# is also feasible to DRCCP (2).

Theorem 2 *Suppose that for any objective upper bound t such that* $t \geq \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top\boldsymbol{x}$, ALSO-X# (5) *is better than* ALSO-X$\underline{\#}$ (6), *i.e.,* $v^{A\#} \leq v^{A\underline{\#}}$.

Proof It is sufficient to show that for a given objective upper bound t, if the lower-level ALSO-X$\underline{\#}$ (6b) obtains a feasible solution to DRCCP (2), then the lower-level ALSO-X# (5b) is also feasible.

Let $(\widehat{\boldsymbol{x}}, \widehat{\beta})$ denote an optimal solution from the lower-level ALSO-X# (5b) and let $(\bar{\boldsymbol{x}}, \bar{\beta})$ denote an optimal solution from the lower-level ALSO-X$\underline{\#}$ (6b). Suppose that $\bar{\boldsymbol{x}}$ is feasible to DRCCP (2), i.e., $\bar{\boldsymbol{x}}$ satisfies (3c). Now let

$$\bar{\beta}^* := \mathrm{VaR}_{1-\varepsilon}\left\{\theta\left\|\boldsymbol{a}_i(\bar{\boldsymbol{x}})\right\|_* + \boldsymbol{a}_i(\bar{\boldsymbol{x}})^\top\tilde{\boldsymbol{\zeta}} - b_i(\bar{\boldsymbol{x}}), \forall i\in[I]\right\} \leq 0.$$

According to theorem 1 in [41] (see, e.g., equation (7) in [41]), then we have that $(\bar{\boldsymbol{x}}, \bar{\beta}^*)$ is another optimal solution to the lower-level ALSO-X$\underline{\#}$ (6b). Since the only difference between the lower-level ALSO-X# (5b) and the lower-level ALSO-X$\underline{\#}$ (6b) is the constraint $\beta \leq 0$, the solution $(\bar{\boldsymbol{x}}, \bar{\beta}^*)$ is also optimal to the lower-level ALSO-X# (5b). That is, for a given objective upper bound t, both lower-level problems have the same optimal value.

In this case, for any optimal solution from the lower-level ALSO-X# (5b), it should also be optimal to the lower-level ALSO-X$\underline{\#}$ (6b). Hence, $(\widehat{\boldsymbol{x}}, \widehat{\beta})$ is also optimal to the lower-level ALSO-X$\underline{\#}$ (6b). Based on theorem 1 in [41] (see, e.g., equation (7) in [41]), we have

$$\widehat{\beta} \geq \mathrm{VaR}_{1-\varepsilon}\left\{\theta\left\|\boldsymbol{a}_i(\widehat{\boldsymbol{x}})\right\|_* + \boldsymbol{a}_i(\widehat{\boldsymbol{x}})^\top\tilde{\boldsymbol{\zeta}} - b_i(\widehat{\boldsymbol{x}}), \forall i\in[I]\right\}.$$

Combining with the condition that $\widehat{\beta} \leq 0$, we have

$$\mathrm{VaR}_{1-\varepsilon}\left\{\theta \left\|\boldsymbol{a}_i(\widehat{\boldsymbol{x}})\right\|_* + \boldsymbol{a}_i(\widehat{\boldsymbol{x}})^\top \tilde{\boldsymbol{\zeta}} - b_i(\widehat{\boldsymbol{x}}), \forall i \in [I]\right\} \leq 0,$$

which implies that $\widehat{\boldsymbol{x}}$ satisfies (3c), i.e., $\widehat{\boldsymbol{x}}$ is also feasible to DRCCP (2). This completes the proof. □

We make the following remarks about Theorem 2:

(i) Note that in Theorem 2, we show that ALSO-X# is better than ALSO-X$\underline{\#}$; and
(ii) One may expect that ALSO-X and ALSO-X$\underline{\#}$ are comparable. In fact, they are not. In Example 1, we can show that ALSO-X$\underline{\#}$ returns a better solution (i.e., $v^{A\underline{\#}} = 0$ and ALSO-X fails to find any feasible solution), while in Example 3, ALSO-X can have a better solution (i.e., $v^A = 1/2 < v^{A\underline{\#}} = 2/3$).

3.3 ALSO-X$\underline{\#}$ is Better Than CVaR Approximation

Recall that the differences between ALSO-X$\underline{\#}$ and CVaR approximation lie in the corresponding upper-level and lower-level problems, where the checking condition in the upper-level CVaR approximation is more restricted than the one in the upper-level ALSO-X$\underline{\#}$, and the lower-level ALSO-X$\underline{\#}$ is a relaxation of the lower-level CVaR approximation. As a result, we show that when CVaR approximation provides a feasible solution to DRCCP (2), any optimal solution of the lower-level ALSO-X$\underline{\#}$ must be feasible to DRCCP (2).

Theorem 3 ALSO-X$\underline{\#}$ (6) *is better than* CVaR *approximation* (4)*, i.e.,* $v^{A\underline{\#}} \leq v^{\mathrm{CVaR}}$.

Proof Notice that for a given objective upper bound t, the lower-level ALSO-X$\underline{\#}$ is a relaxation of the lower-level CVaR approximation. Thus, if the optimal value of the lower-level CVaR approximation is non-positive (i.e., constraint (4c) holds), then the optimal value of the lower-level ALSO-X$\underline{\#}$ must be non-positive, which ensures that any optimal solution is feasible to the DRCCP (2) according to [36]. Therefore, for a given t, ALSO-X$\underline{\#}$ must find a feasible solution to DRCCP (2) if CVaR approximation finds one. This completes the proof. □

We remark that the result in Theorem 3 can be extended to any general ambiguity set, that is, ALSO-X$\underline{\#}$ is better than CVaR approximation under a general ambiguity set. However, this does not hold for Theorem 1, since the worst-case distributions in the lower-level ALSO-X (3b) and the upper-level ALSO-X (3c) may not be the same. Hence, ALSO-X# and ALSO-X are not comparable under the general ambiguity set. Below is an example to illustrate Theorem 3.

Example 4 Consider a single DRCCP under type $\infty-$Wasserstein ambiguity set with $\theta = 1/2$. Assume that the empirical distribution has 3 equiprobable scenarios (i.e., $N = 3$, $\mathbb{P}\{\tilde{\zeta} = \zeta^i\} = 1/N$), risk parameter $\varepsilon = 1/2$, deterministic set $\mathcal{X} = \mathbb{R}_+$, function $a_1(x)^\top\zeta - b_1(x) = x - \zeta$, $\zeta^1 = 5/2$, $\zeta^2 = 3/2$, and $\zeta^3 = 1/2$. In this example, DRCCP (2) resorts to

$$v^* = \min_{x\geq 0}\left\{x\colon \mathbb{I}\left(x \geq 3\right) + \mathbb{I}\left(x \geq 2\right) + \mathbb{I}\left(x \geq 1\right) \geq 2\right\},$$

Its CVaR approximation is

$$v^{\mathrm{CVaR}} = \min_{x\geq 0, \boldsymbol{s}, \beta \leq 0}\left\{x\colon x \geq 3 - s_1, x \geq 2 - s_2, x \geq 1 - s_3, \frac{1}{3}\sum_{i\in[3]} s_i - \frac{\beta}{2} \leq 0, s_i \geq \beta, \forall i \in [3]\right\},$$

and the weak formulation of ALSO-X# (6) can be written as

$$v^{A\underline{\#}} = \min_t \left\{ t\colon \sum_{i\in[3]} \mathbb{I}(s_i^* > 0) \leq 2,\right.$$

$$(x^*, \boldsymbol{s}^*, \beta^*) \in \operatorname*{argmin}_{x \geq 0, \boldsymbol{s}, \beta} \left\{ \frac{1}{3} \sum_{i \in [3]} s_i - \frac{1}{2}\beta \colon \begin{array}{l} x \geq 3 - s_1, x \geq 2 - s_2, x \geq 1 - s_3, \\ x \leq t, s_i \geq \beta, \forall i \in [3] \end{array} \right\}.$$

By the straightforward calculation, we have $v^* = 2$, $v^{\mathrm{CVaR}} = 8/3$, $v^{A\underline{\#}} = 2$. Therefore, in this example, ALSO-X$\underline{\#}$ returns the optimal solution, but CVaR approximation cannot. ⋄

3.4 Summary of Comparisons

Finally, we conclude this section by providing theoretical comparisons among the output objective values of ALSO-X#, ALSO-X$\underline{\#}$, ALSO-X, and CVaR approximation, which are shown in Figure 1.

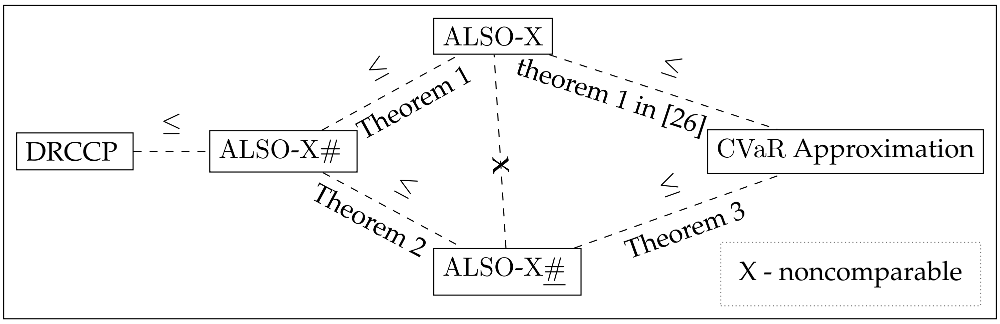


Fig. 1: Summary of Comparisons

4 The Optimal Solution of the Lower-level ALSO-X is Unique

This section investigates conditions under which the lower-level ALSO-X can provide a unique optimal solution, a sufficient condition guaranteeing that ALSO-X# is better than ALSO-X according to Theorem 1. Notably, we prove that the lower-level ALSO-X (3b) admits a unique optimal solution if one of the following conditions hold: (i) empirical data are sampled from continuous nondegenerate distributions and set $\mathcal{X}$ is arbitrary; (ii) a single DRCCP with continuous reference distribution and set $\mathcal{X}$ is convex; (iii) a joint DRCCP with right-hand uncertainty, a continuous reference distribution, and a convex set $\mathcal{X}$; or (iv) a joint DRCCP with left-hand uncertainty and set $\mathcal{X}$ is convex.

4.1 Uniqueness: DRCCPs with an i.i.d. Empirical Reference Distribution Sampling from a Continuous Distribution

In this subsection, we consider the case under which lower-level ALSO-X (3b) can provide a unique solution when the reference distribution is of finite support and is constructed by i.i.d. samples from a continuous nondegenerate distribution. For the given i.i.d. samples of the random parameters $\tilde{\boldsymbol{\zeta}}$, we consider the following ALSO-X:

$$v^A = \min_t \quad t, \tag{7a}$$

$$\text{s.t.} \quad \boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}} \left\{ \frac{1}{N} \sum_{j \in [N]} \max_{i \in [I]} \left[\theta \left\| \boldsymbol{a}_i(\boldsymbol{x}) \right\|_* + \boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\zeta}^j - b_i(\boldsymbol{x}) \right]_+ : \boldsymbol{c}^\top \boldsymbol{x} \leq t, \boldsymbol{x} \in \mathcal{X} \right\}, \tag{7b}$$

$$\sum_{j\in[N]} \mathbb{I}\left\{\max_{i\in[I]} \left[\theta \left\|\boldsymbol{a}_i(\boldsymbol{x}^*)\right\| + \boldsymbol{a}_i(\boldsymbol{x}^*)^\top \boldsymbol{\zeta}^j - b_i(\boldsymbol{x}^*)\right]_+ = 0\right\} \geq N - \lfloor N\varepsilon \rfloor. \tag{7c}$$

In fact, we show that with probability 1, for any objective upper bound t such that $t \geq \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and the optimal value of the lower-level ALSO-X (3b) is positive (i.e., $v^A(t) > 0$), the lower-level ALSO-X (3b) admits a unique optimal solution. Note that if $v^A(t) = 0$, then any optimal solution of the lower-level ALSO-X (3b) is feasible to DRCCP (2). Thus, we focus on the non-trivial case when $v^A(t) > 0$.

Theorem 4 *Suppose (i) for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$ with $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$ and any pair $(i_1, i_2) \in [I] \times [I]$, $\boldsymbol{a}_{i_1}(\boldsymbol{x}_1) \neq \boldsymbol{a}_{i_2}(\boldsymbol{x}_2)$; (ii) the true distribution $\mathbb{P}^*$ of the random parameters $\tilde{\boldsymbol{\zeta}}$ is continuous and nondegenerate; and (iii) $\boldsymbol{\zeta}^1, \boldsymbol{\zeta}^2, \cdots, \boldsymbol{\zeta}^N$ are i.i.d. samples of the random parameters $\tilde{\boldsymbol{\zeta}}$. Then the lower-level* ALSO-X (3b) *admits a unique optimal solution when $t \geq \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and $v^A(t) > 0$.*

Proof We first write the lower-level ALSO-X (7b) as

$$F(\boldsymbol{x}, \mathcal{S}_{\boldsymbol{x}}) \colon = \frac{1}{N} \sum_{j\in[\mathcal{S}_{\boldsymbol{x}}]} \max_{i\in[I]} \left[\theta \left\|\boldsymbol{a}_i(\boldsymbol{x})\right\|_* + \boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\zeta}^j - b_i(\boldsymbol{x})\right], \tag{8}$$

where set $\mathcal{S}_{\boldsymbol{x}}$ is defined as $\mathcal{S}_{\boldsymbol{x}} = \{j \in [N] : \theta \left\|\boldsymbol{a}_i(\boldsymbol{x})\right\|_* + \boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\zeta}^j - b_i(\boldsymbol{x}) \geq 0, \forall i \in [I]\}$. Suppose there exist two different solutions $\boldsymbol{x}_1 \neq \boldsymbol{x}_2 \in \mathcal{X}$ in (8). In this proof, we suppress the notations as $F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_1, \mathcal{S}_{\boldsymbol{x}_1})$ and $F(\boldsymbol{x}_2, \mathcal{S}_2) = F(\boldsymbol{x}_2, \mathcal{S}_{\boldsymbol{x}_2})$. We consider $F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2) > 0$ since $v^A(t) > 0$. We split the remaining proof into three steps based on the deterministic set $\mathcal{X}$.

Step I. Suppose that set $\mathcal{X}$ is compact and discrete. Let i_1^j and i_2^j denote the maximum pieces of the sample $j \in [N]$ in the objective function (8) corresponding to $\boldsymbol{x}_1, \boldsymbol{x}_2$, respectively. There are two cases to discuss.

Case 1). When $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{T} \subseteq [N]$ and $\mathcal{T} \neq \emptyset$, by the definition of $F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2)$, we have

$$\begin{aligned}
&\mathbb{P}^* \left\{\tilde{\boldsymbol{\zeta}} \colon F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2) > 0 \middle| \mathcal{T} \neq \emptyset, i_1^j, i_2^j, \forall j \in [N]\right\} \\
\leq &\mathbb{P}^* \Bigg\{\tilde{\boldsymbol{\zeta}} \colon \sum_{j\in\mathcal{T}} \left[\left[\boldsymbol{a}_{i_1^j}(\boldsymbol{x}_1) - \boldsymbol{a}_{i_2^j}(\boldsymbol{x}_2)\right]^\top \boldsymbol{\zeta}^j\right] \\
&\quad = \sum_{j\in\mathcal{T}} \left[b_{i_1^j}(\boldsymbol{x}_1) - b_{i_2^j}(\boldsymbol{x}_2) + \theta \left[\left\|\boldsymbol{a}_{i_2^j}(\boldsymbol{x})\right\|_* - \left\|\boldsymbol{a}_{i_1^j}(\boldsymbol{x})\right\|_*\right]\right] \Bigg| \mathcal{T} \neq \emptyset, i_1^j, i_2^j, \forall j \in [N]\Bigg\}.
\end{aligned}$$

According to the presumption (i), for any $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$ and any pair $(i_1, i_2) \in [I] \times [I]$, we have $\boldsymbol{a}_{i_1^j}(\boldsymbol{x}_1) \neq \boldsymbol{a}_{i_2^j}(\boldsymbol{x}_2)$ for all $j \in [N]$. Since the random parameters $\tilde{\boldsymbol{\zeta}}$ is continuous and nondegenerate (see, e.g., definition 24.16 in [28]), then

$$\begin{aligned}
\mathbb{P}^* \Bigg\{\tilde{\boldsymbol{\zeta}} \colon &\sum_{j\in\mathcal{T}} \left[\left[\boldsymbol{a}_{i_1^j}(\boldsymbol{x}_1) - \boldsymbol{a}_{i_2^j}(\boldsymbol{x}_2)\right]^\top \boldsymbol{\zeta}^j\right] \\
&= \sum_{j\in\mathcal{T}} \left[b_{i_1^j}(\boldsymbol{x}_1) - b_{i_2^j}(\boldsymbol{x}_2) + \theta \left[\left\|\boldsymbol{a}_{i_2^j}(\boldsymbol{x})\right\|_* - \left\|\boldsymbol{a}_{i_1^j}(\boldsymbol{x})\right\|_*\right]\right] \Bigg| \mathcal{T} \neq \emptyset, i_1^j, i_2^j, \forall j \in [N]\Bigg\} = 0.
\end{aligned}$$

Case 2). When $\mathcal{S}_1 \neq \mathcal{S}_2$, by the definition of $F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2)$, we have

$$\begin{aligned}
&\mathbb{P}^* \left\{\tilde{\boldsymbol{\zeta}} \colon F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2) > 0 \middle| \mathcal{S}_1 \neq \mathcal{S}_2, i_1^j, i_2^j, \forall j \in [N]\right\} \\
\leq &\mathbb{P}^* \Bigg\{\tilde{\boldsymbol{\zeta}} \colon \sum_{j\in\mathcal{S}_1\setminus\mathcal{S}_2} \left[\boldsymbol{a}_{i_1^j}(\boldsymbol{x}_1)^\top \boldsymbol{\zeta}^j\right] - \sum_{j\in\mathcal{S}_2\setminus\mathcal{S}_1} \left[\boldsymbol{a}_{i_2^j}(\boldsymbol{x}_2)^\top \boldsymbol{\zeta}^j\right] + \sum_{j\in\mathcal{S}_1\cap\mathcal{S}_2} \left[\left[\boldsymbol{a}_{i_1^j}(\boldsymbol{x}_1) - \boldsymbol{a}_{i_2^j}(\boldsymbol{x}_2)\right]^\top \boldsymbol{\zeta}^j\right]
\end{aligned}$$

$$= \sum_{j \in \mathcal{T}} \left[b_{i_1^j}(\boldsymbol{x}_1) - b_{i_2^j}(\boldsymbol{x}_2) + \theta \left[\left\| \boldsymbol{a}_{i_2^j}(\boldsymbol{x}) \right\|_* - \left\| \boldsymbol{a}_{i_1^j}(\boldsymbol{x}) \right\|_* \right] \right] \Bigg| \mathcal{S}_1 \neq \mathcal{S}_2, i_1^j, i_2^j, \forall j \in [N] \Bigg\}.$$

Since at least one of the sets $\mathcal{S}_1 \setminus \mathcal{S}_2, \mathcal{S}_2 \setminus \mathcal{S}_1, \mathcal{S}_1 \cap \mathcal{S}_2$ is nonempty, together with the fact that the distribution of random parameters $\tilde{\boldsymbol{\zeta}}$ is continuous and nondegenerate, we have

$$\mathbb{P}^* \Bigg\{ \tilde{\boldsymbol{\zeta}} \colon \sum_{j \in \mathcal{S}_1 \setminus \mathcal{S}_2} \left[\boldsymbol{a}_{i_1^j}(\boldsymbol{x}_1)^\top \boldsymbol{\zeta}^j \right] - \sum_{j \in \mathcal{S}_2 \setminus \mathcal{S}_1} \left[\boldsymbol{a}_{i_2^j}(\boldsymbol{x}_2)^\top \boldsymbol{\zeta}^j \right] + \sum_{j \in \mathcal{S}_1 \cap \mathcal{S}_2} \left[\left[\boldsymbol{a}_{i_1^j}(\boldsymbol{x}_1) - \boldsymbol{a}_{i_2^j}(\boldsymbol{x}_2) \right]^\top \boldsymbol{\zeta}^j \right]$$
$$= \sum_{j \in \mathcal{T}} \left[b_{i_1^j}(\boldsymbol{x}_1) - b_{i_2^j}(\boldsymbol{x}_2) + \theta \left[\left\| \boldsymbol{a}_{i_2^j}(\boldsymbol{x}) \right\|_* - \left\| \boldsymbol{a}_{i_1^j}(\boldsymbol{x}) \right\|_* \right] \right] \Bigg| \mathcal{S}_1 \neq \mathcal{S}_2, i_1^j, i_2^j, \forall j \in [N] \Bigg\} = 0.$$

Combining these two cases, for a given $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$, we have

$$\begin{aligned}
&\mathbb{P}^* \{ \tilde{\boldsymbol{\zeta}} \colon F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2) > 0 \} \\
=& \sum_{\mathcal{T} \subseteq [N], i_1^j \in [I], i_2^j \in [I], \forall j \in \mathcal{T}} \mathbb{P}^* \left\{ \mathcal{T} \neq \emptyset, i_1^j, i_2^j, \forall j \in \mathcal{T} \right\} \mathbb{P}^* \left\{ \tilde{\boldsymbol{\zeta}} \colon F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2) > 0 \Big| \mathcal{T} \neq \emptyset, i_1^j, i_2^j, \forall j \in [N] \right\} \\
&+ \sum_{\mathcal{S}_1, \mathcal{S}_2 \subseteq [N], \mathcal{S}_1 \neq \mathcal{S}_2, i_1^j \in, i_2^j \in \mathcal{S}_2, \forall j \in \mathcal{S}_1 \cup \mathcal{S}_2} \mathbb{P}^* \left\{ \mathcal{S}_1 \neq \mathcal{S}_2, i_1^j, i_2^j, \forall j \in \mathcal{S}_1 \cup \mathcal{S}_2 \right\} \\
&\mathbb{P}^* \left\{ \tilde{\boldsymbol{\zeta}} \colon F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2) > 0 \Big| \mathcal{S}_1 \neq \mathcal{S}_2, i_1^j, i_2^j, \forall j \in [N] \right\} \\
=& 0.
\end{aligned}$$

Hence, for any $\boldsymbol{x}_1 \neq \boldsymbol{x}_2 \in \mathcal{X}$, we further have

$$\mathbb{P}^* \left\{ \tilde{\boldsymbol{\zeta}} \colon \bigcup_{\substack{\boldsymbol{x}_1 \in \mathcal{X}, \boldsymbol{x}_2 \in \mathcal{X}, \\ \boldsymbol{x}_1 \neq \boldsymbol{x}_2}} F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2) > 0 \right\} = 0.$$

Therefore, we show that when set $\mathcal{X}$ is discrete and compact, there exists a unique optimal solution in the lower-level (8) with probability 1.

Step II. Suppose that set $\mathcal{X}$ is compact but may not be discrete. Suppose set $\mathcal{X} \subseteq [-M, M]^n$. Then for some small $\nu > 0$, by discretization, we have for any $\boldsymbol{x} \in \mathcal{X}$, there exists $\boldsymbol{y} \in \mathcal{X}^\nu$, such that $\|\boldsymbol{x} - \boldsymbol{y}\|_\infty \leq \nu$ and $|\mathcal{X}^\nu| \leq |2M/\nu|^n$. Instead of optimizing over $\mathcal{X}$, we consider optimizing over $\mathcal{X}^\nu$ in (8). Here, we choose ν as $\nu(\tau) = 2M/2^\tau, \tau \in \mathbb{N}$. Following the similar procedures in Step I, we assume that there exist two different solutions $\boldsymbol{x}_1(\nu(\tau)) \neq \boldsymbol{x}_2(\nu(\tau))$ such that $\boldsymbol{x}_1(\nu(\tau)), \boldsymbol{x}_2(\nu(\tau)) \in \mathcal{X}^\nu(\tau), F(\boldsymbol{x}_1(\nu(\tau)), \mathcal{S}_1) = F(\boldsymbol{x}_2(\nu(\tau)), \mathcal{S}_2) > 0$. For any $\tau \in \mathbb{N}$, we have

$$\mathbb{P}^* \left\{ \tilde{\boldsymbol{\zeta}} \colon \bigcup_{\substack{\boldsymbol{x}_1(\nu(\tau)) \in \mathcal{X}^\nu(\tau), \boldsymbol{x}_2(\nu(\tau)) \in \mathcal{X}^\nu(\tau), \\ \boldsymbol{x}_1(\nu(\tau)) \neq \boldsymbol{x}_2(\nu(\tau))}} F(\boldsymbol{x}_1(\nu(\tau)), \mathcal{S}_1) = F(\boldsymbol{x}_2(\nu(\tau)), \mathcal{S}_2) > 0 \right\} = 0.$$

When τ increases, the number of feasible solutions increases. That is, the measurable sequence

$$\bigcup_{\substack{\boldsymbol{x}_1(\nu(\tau)) \in \mathcal{X}^\nu(\tau), \boldsymbol{x}_2(\nu(\tau)) \in \mathcal{X}^\nu(\tau), \\ \boldsymbol{x}_1(\nu(\tau)) \neq \boldsymbol{x}_2(\nu(\tau))}} \{ F(\boldsymbol{x}_1(\nu(\tau)), \mathcal{S}_1) = F(\boldsymbol{x}_2(\nu(\tau)), \mathcal{S}_2) > 0 \}$$

is monotone nondecreasing as τ increases. Following the Monotone Convergence Theorem for sequences of measurable sets (see, e.g., theorem 1.26 in [52]), when $\tau \to \infty$, the limit of this measurable sequence exists. Thus, we have

$$\mathbb{P}^* \left\{ \tilde{\boldsymbol{\zeta}} \colon \lim_{\tau \to \infty} \bigcup_{\substack{\boldsymbol{x}_1(\nu(\tau)) \in \mathcal{X}^\nu(\tau), \boldsymbol{x}_2(\nu(\tau)) \in \mathcal{X}^\nu(\tau), \\ \boldsymbol{x}_1(\nu(\tau)) \neq \boldsymbol{x}_2(\nu(\tau))}} F(\boldsymbol{x}_1(\nu(\tau)), \mathcal{S}_1) = F(\boldsymbol{x}_2(\nu(\tau)), \mathcal{S}_2) > 0 \right\}$$
$$= \lim_{\tau \to \infty} \mathbb{P}^* \left\{ \tilde{\boldsymbol{\zeta}} \colon \bigcup_{\substack{\boldsymbol{x}_1(\nu(\tau)) \in \mathcal{X}^\nu(\tau), \boldsymbol{x}_2(\nu(\tau)) \in \mathcal{X}^\nu(\tau), \\ \boldsymbol{x}_1(\nu(\tau)) \neq \boldsymbol{x}_2(\nu(\tau))}} F(\boldsymbol{x}_1(\nu(\tau)), \mathcal{S}_1) = F(\boldsymbol{x}_2(\nu(\tau)), \mathcal{S}_2) > 0 \right\} = 0.$$

Thus, when set $\mathcal{X}$ is compact but not discrete, there exists a unique solution of (8) with probability 1.

Step III. The result holds when set $\mathcal{X}$ is not compact. Let $\widehat{\mathcal{X}}_r = \mathcal{X} \cap \mathcal{B}(0, r)$ with $r > 0$. By definition, $\widehat{\mathcal{X}}_r$ is compact. Then following the similar procedures in Step I and Step II, there exist two different solutions $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$ such that $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \widehat{\mathcal{X}}_r, F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2) > 0$. For any $r > 0$, since set $\widehat{\mathcal{X}}_r$ is bounded, according to Step II, we have

$$\mathbb{P}^* \left\{ \tilde{\boldsymbol{\zeta}} \colon \bigcup_{\substack{\boldsymbol{x}_1 \in \widehat{\mathcal{X}}_r, \boldsymbol{x}_2 \in \widehat{\mathcal{X}}_r, \\ \boldsymbol{x}_1 \neq \boldsymbol{x}_2}} F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2) > 0 \right\} = 0.$$

Due to Monotone Convergence Theorem for sequences of measurable sets (see, e.g., theorem 1.26 in [52]), we have

$$\mathbb{P}^* \left\{ \tilde{\boldsymbol{\zeta}} \colon \lim_{r \to \infty} \bigcup_{\substack{\boldsymbol{x}_1 \in \widehat{\mathcal{X}}_r, \boldsymbol{x}_2 \in \widehat{\mathcal{X}}_r, \\ \boldsymbol{x}_1 \neq \boldsymbol{x}_2}} F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2) > 0 \right\}$$
$$= \lim_{r \to \infty} \mathbb{P}^* \left\{ \tilde{\boldsymbol{\zeta}} \colon \bigcup_{\substack{\boldsymbol{x}_1 \in \widehat{\mathcal{X}}_r, \boldsymbol{x}_2 \in \widehat{\mathcal{X}}_r, \\ \boldsymbol{x}_1 \neq \boldsymbol{x}_2}} F(\boldsymbol{x}_1, \mathcal{S}_1) = F(\boldsymbol{x}_2, \mathcal{S}_2) > 0 \right\} = 0.$$

Therefore, when set $\mathcal{X}$ is not compact, there exists a unique solution of (8) with probability 1. This completes the proof. □

We make the following remarks on Theorem 4:

(i) The proof shows that any objective value of the lower-level ALSO-X is unique; and
(ii) The key of the proof is to exploit the properties of the linear uncertain constraints and continuous nondegenerate distribution. Relaxing any of them, the result in Theorem 4 may not hold.

Sampling from a continuous distribution helps us find a unique solution of the lower-level ALSO-X (3b). However, if the conditions in Theorem 4 were not met, there might not be a unique optimal solution. For example, when the true distribution of random parameters $\tilde{\boldsymbol{\xi}}$ is of discrete support, the lower-level ALSO-X may not have a unique solution, i.e., in Example 3, when $t = 1/2$, the lower-level ALSO-X provides two optimal solutions, of which one is feasible to DRCCP (2) and another one is not.

We use the following example to show that the lower-level ALSO-X (3b) does not admit a unique solution either when the reference distribution is continuous.

Example 5 Consider a single DRCCP under type $\infty-$Wasserstein ambiguity set with a Gaussian distribution $\tilde{\boldsymbol{\zeta}} \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$, and $\theta = 1/2$) with $n = 3, \bar{\boldsymbol{\mu}} = [1, 1, 1]^\top$, $\bar{\boldsymbol{\Sigma}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, risk parameter $\varepsilon = 0.40$, set $\mathcal{X} = \{0, 1\}^3$ and function $\boldsymbol{a}_1(\boldsymbol{x})^\top \boldsymbol{\zeta} - b_1(\boldsymbol{x}) = -2 + \boldsymbol{x}^\top \boldsymbol{\zeta}$. In this example, the lower-level ALSO-X (3b) can be written as

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}} \left\{ \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left[\theta \left\| \boldsymbol{x} \right\|_* + \tilde{\zeta}_1 x_1 + \tilde{\zeta}_2 x_2 + \tilde{\zeta}_3 x_3 - 2 \right]_+ : \, - x_1 - x_2 - x_3 \leq t, \boldsymbol{x} \in \{0, 1\}^3 \right\},$$

where the dual norm is $\|\cdot\|_2$. Let $t = -2$. A simple calculation shows that there are three optimal solutions from the lower-level ALSO-X (3b) with the same positive objective value, i.e., $(\boldsymbol{x}^1)^* = (1, 1, 0)^\top$, $(\boldsymbol{x}^2)^* = (1, 0, 1)^\top$, $(\boldsymbol{x}^3)^* = (0, 1, 1)^\top$. Therefore, in this case, there is no unique solution from the lower-level ALSO-X. $\diamond$

This motivates us to restrict set $\mathcal{X}$ to be convex when the reference distribution is continuous in the next subsections.

4.2 Uniqueness: Single DRCCPs with Continuous Reference Distributions

We consider the case when the reference distribution is continuous. For a single DRCCP (2), i.e., $I = 1$, when the affine mappings are $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$, $b_1(\boldsymbol{x}) = b_1$, the random parameters $\tilde{\boldsymbol{\zeta}}$ is continuous and the deterministic set $\mathcal{X}$ is convex, the lower-level ALSO-X (3b) is equivalent to

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}} \left\{ \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left[\theta \left\| \boldsymbol{x} \right\|_* + \boldsymbol{x}^\top \tilde{\boldsymbol{\zeta}} - b_1 \right]_+ : \boldsymbol{c}^\top \boldsymbol{x} \leq t, \boldsymbol{x} \in \mathcal{X} \right\}.$$

Recall that $\widehat{F}(\boldsymbol{x})$ denotes the objective function in the lower-level ALSO-X (3b). In this case, we have

$$\widehat{F}(\boldsymbol{x}) = \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left[\theta \left\| \boldsymbol{x} \right\|_* + \boldsymbol{x}^\top \tilde{\boldsymbol{\zeta}} - b_1 \right]_+.$$

Theorem 5 *Suppose that in a single DRCCP* (2), *the deterministic set $\mathcal{X}$ is convex, the reference distribution $\mathbb{P}_{\tilde{\zeta}}$ is continuous and nondegenerate with support $\mathbb{R}^n$, $\|\cdot\|_* = \|\cdot\|_p$ with $p \in (1, \infty)$ and affine mappings $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$ and $b_1(\boldsymbol{x}) = b_1$. Then the lower-level* ALSO-X (3b) *admits a unique optimal solution when $t \geq \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and $v^A(t) > 0$.*

Proof See Appendix A.1. □

Note that our analysis in Theorem 5 shows that the uniqueness of the lower-level ALSO-X applies to any general continuous nondegenerate distribution. Then the following corollary shows that the lower-level ALSO-X (3b) returns a unique optimal solution when we know the upper and lower bounds of the support Ξ.

Corollary 1 *Suppose that in a single DRCCP* (2), *the deterministic set $\mathcal{X}$ is convex, $\|\cdot\|_* = \|\cdot\|_p$ with $p \in (1, \infty)$ and affine mappings $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$ and $b_1(\boldsymbol{x}) = b_1$, and the reference distribution $\mathbb{P}_{\tilde{\zeta}}$ is continuous and nondegenerate with a closed convex support Ξ such that*

$$\exists \boldsymbol{\zeta} \in \Xi \colon \min_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{c}^\top \boldsymbol{x} \leq t} \left\{ \theta \|\boldsymbol{x}\|_* + \boldsymbol{x}^\top \boldsymbol{\zeta} \right\} > b_1, \text{ and } \exists \boldsymbol{\zeta} \in \Xi \colon \max_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{c}^\top \boldsymbol{x} \leq t} \left\{ \theta \|\boldsymbol{x}\|_* + \boldsymbol{x}^\top \boldsymbol{\zeta} \right\} < b_1.$$

Then the lower-level ALSO-X (3b) *admits a unique optimal solution when $t \geq \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and $v^A(t) > 0$.*

Proof Using the fact that in Part (i) of the proof of Theorem 5, we also have $\int_{\partial \mu(\boldsymbol{x}^*)} (\boldsymbol{y}^\top \partial(\theta \|\boldsymbol{x}\|_*)/\partial \boldsymbol{x} + \boldsymbol{y}^\top \boldsymbol{\zeta})^2 \mathbb{P}(d\boldsymbol{\zeta}) > 0$ according to the presumption, the proof is almost identical to that of Theorem 5 and is thus omitted. □

We remark that the result in Theorem 5 can also be generalized to $\|\cdot\|_* = \|\cdot\|_p$ with $p \in \{1, \infty\}$. Due to the page limit, we refer interested readers to Appendix A.2 for the proof.

4.3 Uniqueness: Joint DRCCPs with a Continuous Reference Distribution

In this subsection, we consider a joint DRCCP with right-hand uncertainty and a continuous reference distribution. In particular, we assume that $I = n$, the uncertainty constraint is $\boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\zeta} - b_i(\boldsymbol{x}) = \zeta_i - x_i$, and the random parameter $\tilde{\zeta}_i$ is continuous for each $i \in [n]$. That is, we consider the following DRCCP:

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \inf_{\mathbb{P}\in\mathcal{P}_\infty} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} \colon \tilde{\xi}_i \le x_i, \forall i \in [n] \right\} \ge 1 - \varepsilon \right\},$$

that is,

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \mathbb{P}\left\{ \tilde{\boldsymbol{\zeta}} \colon \tilde{\zeta}_i + \theta \le x_i, \forall i \in [n] \right\} \ge 1 - \varepsilon \right\}. \tag{9}$$

In this case, the lower-level ALSO-X (3b) is

$$v^A(t) = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \mathbb{E}_{\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}} \left[\max_{i\in[n]} \left\{ \tilde{\zeta}_i + \theta - x_i \right\}_+ \right] : \boldsymbol{c}^\top \boldsymbol{x} \le t \right\}.$$

Our proof idea is to show the positive definiteness of the Hessian of the objective function in the lower-level ALSO-X (3b) to prove the uniqueness.

Theorem 6 *Suppose that in a joint DRCCP* (9)*, the deterministic set $\mathcal{X}$ is convex, and the reference distribution $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ is continuous with support $\mathbb{R}^n$. Then the lower-level* ALSO-X (3b) *admits a unique optimal solution when $t \ge \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and $v^A(t) > 0$.*

Proof Suppose that there exists an optimal solution $\boldsymbol{x}^*$ to the lower-level ALSO-X (3b). We split the proof into two steps to show that $\boldsymbol{x}^*$ is the unique solution.
Step I. We first provide the closed-form expression of the lower-level ALSO-X (3b) and its first-order and second-order derivatives, where the objective function in the lower-level ALSO-X (3b) is $\widehat{F}(\boldsymbol{x}^*) = \max_{i\in[n]} \{\zeta_i + \theta - x_i^*\}_+$. According to the continuity of function $f(\tau) = \max\{\tau, 0\}$ and theorem 1 in [42], we can interchange the subdifferential operator and expectation, the first-order derivative for each $i \in [n]$ of $\widehat{F}(\boldsymbol{x}^*)$ is

$$\frac{\partial \widehat{F}(\boldsymbol{x}^*)}{\partial x_i} = -\int_{\mu_i(\boldsymbol{x}^*)} \mathbb{P}(d\boldsymbol{\zeta}),$$

where $\mu_i(\boldsymbol{x}) = \{\boldsymbol{\zeta} : \zeta_i + \theta - x_i \ge 0, \zeta_i + \theta - x_i \ge \zeta_j + \theta - x_j, \forall j \in [n] \setminus \{i\}\}$. Let us take its second derivative, i.e., for each $i \in [n]$, the diagonal entry of the Hessian matrix $H_{\widehat{F}}(\boldsymbol{x}^*)$ is

$$\frac{\partial^2 \widehat{F}(\boldsymbol{x}^*)}{\partial x_i^2} = \int_{\partial\mu_{i0}(\boldsymbol{x}^*)} \mathbb{P}(d\boldsymbol{\zeta}) + \sum_{\tau\in[n]\setminus\{i\}} \int_{\partial\mu_{i\tau}(\boldsymbol{x}^*)} \mathbb{P}(d\boldsymbol{\zeta}),$$

where

$$\partial\mu_{i0}(\boldsymbol{x}) = \{\boldsymbol{\zeta} : \zeta_i + \theta - x_i = 0 \ge \zeta_j + \theta - x_j, \forall j \in [n] \setminus \{i\}\},$$

and

$$\partial\mu_{i\tau}(\boldsymbol{x}) = \{\boldsymbol{\zeta} : \zeta_i + \theta - x_i \ge 0, \zeta_i + \theta - x_i = \zeta_\tau + \theta - x_\tau \ge \zeta_j + \theta - x_j, \forall j \in [n] \setminus \{i\}\}.$$

For each $\tau \in [n] \setminus \{i\}$, the off-diagonal entry of the Hessian matrix $H_{\widehat{F}}(\boldsymbol{x}^*)$ is

$$\frac{\partial^2 \widehat{F}(\boldsymbol{x}^*)}{\partial x_i \partial x_\tau} = -\int_{\partial\mu_{i\tau}(\boldsymbol{x}^*)} \mathbb{P}(d\boldsymbol{\zeta}).$$

Step II. Next, we show that the Hessian matrix $H_{\widehat{F}}(\boldsymbol{x}^*)$ is strictly diagonally dominant. We split the following proof into two parts: (i) $\int_{\partial\mu_{i0}(\boldsymbol{x}^*)} \mathbb{P}(d\boldsymbol{\zeta}) > 0$ for each $i \in [n]$; and (ii) $\int_{\partial\mu_{i\tau}(\boldsymbol{x}^*)} \mathbb{P}(d\boldsymbol{\zeta}) > 0$ for each $i \in [n]$ and $\tau \in [n] \setminus \{i\}$.
Part (i). The fact that $\int_{\partial\mu_{i0}(\boldsymbol{x}^*)} \mathbb{P}(d\boldsymbol{\zeta}) > 0$ for each $i \in [n]$ is because set $\partial\mu_{i0}(\boldsymbol{x}^*)$ has a dimension of $n-1$ and set $\partial\mu_{i0}(\boldsymbol{x}^*)$ has a nonempty relative interior.
Part (ii). The fact that $\int_{\partial\mu_{i\tau}(\boldsymbol{x}^*)} \mathbb{P}(d\boldsymbol{\zeta}) > 0$ for each $i \in [n]$ and $\tau \in [n] \setminus \{i\}$ is because set $\partial\mu_{i\tau}(\boldsymbol{x}^*)$ has a dimension of $n-1$ and set $\partial\mu_{i\tau}(\boldsymbol{x}^*)$ has a nonempty relative interior.

Thus, the Hessian matrix $H_{\widehat{F}}(\boldsymbol{x}^*)$ is strictly diagonally dominant, i.e., the following two conditions satisfied: (i) for each $i \in [n]$ and $\tau \in [n]\setminus\{i\}$, $\partial^2 \widehat{F}(\boldsymbol{x}^*)/\partial x_i \partial x_\tau < 0$; and (ii) for each $i \in [n]$, $\partial^2 \widehat{F}(\boldsymbol{x}^*)/\partial x_i^2 + \sum_{\tau\in[n]\setminus\{j\}} \partial^2 \widehat{F}(\boldsymbol{x}^*)/\partial x_i \partial x_\tau > 0$. According to Gershgorin circle theorem (see, e.g., theorem 6.1.10 in [24]), the Hessian matrix $H_{\widehat{F}}(\boldsymbol{x}^*)$ is positive definite. Therefore, the optimal solution $\boldsymbol{x}^*$ is unique. □

The following corollary shows that the lower-level ALSO-X (3b) returns a unique optimal solution when Ξ is closed and convex with mild conditions.

Corollary 2 *Suppose support Ξ of $\tilde{\boldsymbol{\zeta}}$ is closed and convex. When $t \geq \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and $v^A(t) > 0$, for any $\boldsymbol{x} \in \mathcal{X} \cap \{\boldsymbol{c}^\top \boldsymbol{x} \leq t\}$ with a positive lower-level* ALSO-X *objective value, the set*

$$\partial\mu_{i0}(\boldsymbol{x}) = \{\boldsymbol{\zeta} \in \Xi : \zeta_i + \theta - x_i = 0 \geq \zeta_j + \theta - x_j, \forall j \in [n] \setminus \{i\}\}$$

has a dimension of $n-1$. In a joint DRCCP (9), when the deterministic set $\mathcal{X}$ is convex and the reference distribution $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ is continuous with support Ξ. Then the lower-level ALSO-X (3b) *admits a unique optimal solution when $t \geq \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and $v^A(t) > 0$.*

Proof Using the fact that similar to part (ii) in the proof of Theorem 6, we also have $\int_{\partial\mu_{i0}(\boldsymbol{x}^*)} \mathbb{P}(d\boldsymbol{\zeta}) > 0$ for each $i \in [n]$ according to the presumption, the proof is almost identical to that of Theorem 6 and is thus omitted. □

4.4 Uniqueness: Joint DRCCPs with Left-hand Side Uncertainty

We consider the case when the reference distribution is continuous. For a joint DRCCP (2) with left-hand side uncertainty and knapsack constraints, we assume that the affine mappings are $\boldsymbol{a}_i(\boldsymbol{x}) = \boldsymbol{x}$, $b_i(\boldsymbol{x}) = b_i$ for each $i \in [I]$, the random parameters $\tilde{\boldsymbol{\zeta}}$ is continuous with $\boldsymbol{\zeta} := [\boldsymbol{\zeta}_1, \cdots, \boldsymbol{\zeta}_I]$ such that $\boldsymbol{\zeta}_i$ and $\boldsymbol{\zeta}_j$ do not overlap for each $i \neq j$, and the deterministic set $\mathcal{X}$ is convex. That is, we consider the following DRCCP:

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \inf_{\mathbb{P}\in\mathcal{P}_\infty} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} \colon \boldsymbol{x}^\top \tilde{\boldsymbol{\xi}}_i \leq b_i, \forall i \in [I] \right\} \geq 1 - \varepsilon \right\},$$

that is,

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \mathbb{P}\left\{ \tilde{\boldsymbol{\zeta}} \colon \theta \left\|\boldsymbol{x}\right\|_* + \boldsymbol{x}^\top \tilde{\boldsymbol{\zeta}}_i \leq b_i, \forall i \in [I] \right\} \geq 1 - \varepsilon \right\}. \tag{10}$$

In this case, the lower-level ALSO-X (3b) is

$$v^A(t) = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left[\max_{i\in[I]} \left\{ \theta \left\|\boldsymbol{x}\right\|_* + \boldsymbol{x}^\top \tilde{\boldsymbol{\zeta}}_i - b_i \right\}_+ \right] : \boldsymbol{c}^\top \boldsymbol{x} \leq t \right\}.$$

Recall that $\widehat{F}(\boldsymbol{x})$ denotes the objective function in the lower-level ALSO-X (3b). Under this circumstance, we have

$$\widehat{F}(\boldsymbol{x}) = \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left[\max_{i\in[I]} \left\{ \theta \left\|\boldsymbol{x}\right\|_* + \boldsymbol{x}^\top \tilde{\boldsymbol{\zeta}}_i - b_i \right\}_+ \right].$$

Theorem 7 *Suppose that in a joint DRCCP* (10), *the deterministic set $\mathcal{X}$ is convex, the reference distribution $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ is continuous and nondegenerate with support $\mathbb{R}^n$, $\|\cdot\|_* = \|\cdot\|_p$ with $p \in (1, \infty)$, affine mappings $\boldsymbol{a}_i(\boldsymbol{x}) = \boldsymbol{x}$ and $b_i(\boldsymbol{x}) = b_i$ for each $i \in [I]$, and the random parameters $\tilde{\boldsymbol{\zeta}}$ is continuous with $\boldsymbol{\zeta} := [\boldsymbol{\zeta}_1, \cdots, \boldsymbol{\zeta}_I]$ such that $\boldsymbol{\zeta}_i$ and $\boldsymbol{\zeta}_j$ do not overlap for each $i \neq j$. Then the lower-level* ALSO-X (3b) *admits a unique optimal solution when $t \geq \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and $v^A(t) > 0$.*

Proof The first-order derivative of $\widehat{F}(\boldsymbol{x})$ is

$$\widehat{F}'(\boldsymbol{x}) = \frac{\partial \widehat{F}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \sum_{i \in [I]} \int_{\mu_i(\boldsymbol{x})} \frac{\partial}{\partial \boldsymbol{x}} \left[\theta \left\|\boldsymbol{x}\right\|_* + \boldsymbol{x}^\top \boldsymbol{\zeta}_i\right] \mathbb{P}(d\boldsymbol{\zeta}),$$

where for each $i \in [I]$, $\mu_i(\boldsymbol{x}) = \{\boldsymbol{\zeta} : \theta \|\boldsymbol{x}\|_* + \boldsymbol{x}^\top \boldsymbol{\zeta}_i \geq b_i, \boldsymbol{x}^\top \boldsymbol{\zeta}_i - b_i \geq \boldsymbol{x}^\top \boldsymbol{\zeta}_j - b_j, \forall j \in [I] \setminus \{i\}\}$; and the Hessian of $\widehat{F}(\boldsymbol{x})$ is

$$\begin{aligned} H_{\widehat{F}}(\boldsymbol{x}) = \sum_{i \in [I]} \Bigg[& \frac{1}{\|\boldsymbol{x}\|_2} \int_{\mu_i(\boldsymbol{x}) \cap \{\theta \|\boldsymbol{x}\|_* + \boldsymbol{x}^\top \boldsymbol{\zeta}_i = b_i\}} \left(\frac{\partial \theta \left\|\boldsymbol{x}\right\|_*}{\partial \boldsymbol{x}} + \boldsymbol{\zeta}_i \right) \left(\frac{\partial \theta \left\|\boldsymbol{x}\right\|_*}{\partial \boldsymbol{x}} + \boldsymbol{\zeta}_i \right)^\top \mathbb{P}(d\boldsymbol{\zeta}) \\ & + \frac{1}{\|\boldsymbol{x}\|_2} \sum_{\tau \in [I] \setminus \{i\}} \int_{\mu_i(\boldsymbol{x}) \cap \{\boldsymbol{x}^\top \boldsymbol{\zeta}_i - b_i = \boldsymbol{x}^\top \boldsymbol{\zeta}_\tau - b_\tau\}} (\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_\tau)(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_\tau)^\top \mathbb{P}(d\boldsymbol{\zeta}) + \theta \int_{\mu_i(\boldsymbol{x})} \frac{\partial^2 \|\boldsymbol{x}\|_*}{\partial \boldsymbol{x}^2} \mathbb{P}(d\boldsymbol{\zeta}) \Bigg]. \end{aligned}$$

The remaining proof is similar to that of Theorem 5 and is thus omitted for brevity. □

Similarly, when the upper and lower bounds of the support $\varXi$ are accessible, the following corollary shows that the lower-level ALSO-X (3b) returns a unique optimal solution.

Corollary 3 *Suppose support $\varXi$ of $\tilde{\boldsymbol{\zeta}}$ is closed and convex. When $t \geq \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and $v^A(t) > 0$, for any $\boldsymbol{x} \in \mathcal{X} \cap \{\boldsymbol{c}^\top \boldsymbol{x} \leq t\}$ with a positive lower-level* ALSO-X *objective value, the set*

$$\partial \mu_i(\boldsymbol{x}) = \left\{ \boldsymbol{\zeta} \in \varXi : \theta \left\|\boldsymbol{x}\right\|_* + \boldsymbol{x}^\top \boldsymbol{\zeta}_i - b_i = 0 \geq \theta \left\|\boldsymbol{x}\right\|_* + \boldsymbol{x}^\top \boldsymbol{\zeta}_j - b_j, \forall j \in [n] \setminus \{i\} \right\}$$

has a dimension of $n - 1$. In a joint DRCCP (10), *when the deterministic set $\mathcal{X}$ is convex and the reference distribution $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ is continuous support $\varXi$. Then the lower-level* ALSO-X (3b) *admits a unique optimal solution when $t \geq \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and $v^A(t) > 0$.*

5 Exactness: ALSO-X# Provides an Optimal Solution to a DRCCP

In this section, we provide sufficient conditions under which ALSO-X# provides an optimal solution to a DRCCP under type $\infty-$Wasserstein ambiguity set. According to Theorem 2, $v^{A\#} \leq v^{A\underline{\#}}$. Thus, for ease of analysis, we focus on ALSO-X$\underline{\#}$ (6). To begin with, we recast DRCCP (2) and ALSO-X$\underline{\#}$ (6) in the following forms:

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \widehat{G}_\theta \left(\boldsymbol{x}^\top \boldsymbol{h} \right) \leq 0 \right\}; \tag{11}$$

and

$$\begin{aligned} v^{A\underline{\#}} = \min_t \quad & t, && \text{(12a)} \\ \text{s.t.} \quad & \boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X}} \left\{ \overline{F}_\theta \left(\boldsymbol{x}^\top \boldsymbol{h} \right) : \boldsymbol{c}^\top \boldsymbol{x} \leq t \right\}, && \text{(12b)} \\ & \widehat{G}_\theta \left((\boldsymbol{x}^*)^\top \boldsymbol{h} \right) \leq 0. && \text{(12c)} \end{aligned}$$

We see that if both functions $\widehat{G}_\theta(\cdot)$ and $\overline{F}_\theta(\cdot)$ are monotone nondecreasing, then ALSO-X$\underline{\#}$ can find an optimal solution to DRCCP. So is ALSO-X# according to Theorem 2.

Theorem 8 *Suppose that in DRCCP* (11) *and* ALSO-X$\underline{\#}$ (12)*, both functions $\widehat{G}_\theta(\cdot)$ and $\overline{F}_\theta(\cdot)$ are monotone nondecreasing. Then* ALSO-X# *is exact.*

Proof Let v_1, v_2 be the optimal values of DRCCP (11) and ALSO-X$\underline{\#}$ (12), respectively. Since ALSO-X$\underline{\#}$ (12) is a conservative approximation, we must have $v_1 \le v_2$. Then it remains to show that $v_2 \le v_1$.

Let $\boldsymbol{x}^*$ be an optimal solution of DRCCP (11) and $t^* = \boldsymbol{c}^\top \boldsymbol{x}^*$. Plug t^* into the lower-level ALSO-X$\underline{\#}$ and let $\widehat{\boldsymbol{x}}$ be its optimal solution, that is,

$$\widehat{\boldsymbol{x}} \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} \left\{ \overline{F}_\theta \left(\boldsymbol{x}^\top \boldsymbol{h}\right) : \boldsymbol{c}^\top \boldsymbol{x} \le t^* \right\}.$$

Since $\boldsymbol{x}^*$ is feasible to the lower-level ALSO-X$\underline{\#}$ with $t^* = \boldsymbol{c}^\top \boldsymbol{x}^*$, we must have $\overline{F}_\theta(\widehat{\boldsymbol{x}}^\top \boldsymbol{h}) \le \overline{F}_\theta((\boldsymbol{x}^*)^\top \boldsymbol{h})$. According to the monotonicity assumption of the function $\widehat{G}_\theta(\boldsymbol{x}^\top \boldsymbol{h})$, we further have $\widehat{G}_\theta(\widehat{\boldsymbol{x}}^\top \boldsymbol{h}) \le \widehat{G}_\theta((\boldsymbol{x}^*)^\top \boldsymbol{h})$, which implies that $\widehat{G}_\theta(\widehat{\boldsymbol{x}}^\top \boldsymbol{h}) \le 0$. Hence, we have $v_2 \le t^* = v_1$. That is, ALSO-X$\underline{\#}$ is exact.

According to Theorem 2, ALSO-X# is less conservative than ALSO-X$\underline{\#}$. Thus, ALSO-X# is also exact. This completes the proof. □

Next, we identify three special families of DRCCPs satisfying the conditions in Theorem 8, namely, single DRCCPs with elliptical, multinomial, and finite-support reference distributions, respectively.

5.1 Special Case I: Single DRCCPs with Elliptical Reference Distributions

We consider a single DRCCP when the reference distribution is elliptical. Note that an elliptical distribution $\mathbb{P}_{\mathrm{E}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \widehat{g})$ is described by three parameters, a location parameter $\boldsymbol{\mu}$, a positive semi-definite matrix $\boldsymbol{\Sigma}$, and a generating function $\widehat{g}$, and its probability density function $\widehat{f}$ has the following form:

$$\widehat{f}(\boldsymbol{x}) = \bar{k} \cdot \widehat{g}\left(\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

with a positive normalization scalar $\bar{k}$. The probability density function of the standard univariate elliptical distribution $\mathbb{P}_{\mathrm{E}}(0,1,\widehat{g})$ is $\varphi(z) = \bar{k}\widehat{g}(z^2/2)$, and the corresponding cumulative distribution function is $\Phi(\tau) = \int_{-\infty}^{\tau} \bar{k}\widehat{g}(z^2/2)dz$. For the single DRCCP (2), i.e., $I = 1$, when the affine mappings are $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$, $b_1(\boldsymbol{x}) = b_1$, the random parameters $\tilde{\boldsymbol{\zeta}}$ follow a joint elliptical distribution with $\tilde{\boldsymbol{\zeta}} \sim \mathbb{P}_{\mathrm{E}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \widehat{g})$, and the norm defining the Wasserstein distance is the generalized Mahalanobis norm associated with the matrix $\boldsymbol{\Sigma}$, i.e., $\|\boldsymbol{y}\| = \sqrt{\boldsymbol{y}^\top \boldsymbol{\Sigma}^\dagger \boldsymbol{y}}$, for some $\boldsymbol{y} \in \mathbb{R}^n$, where $\boldsymbol{\Sigma}^\dagger$ is the pseudo-inverse. According to the reformulations in proposition 10 of [26], DRCCP (2) resorts to

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \boldsymbol{\mu}^\top \boldsymbol{x} + \left(\Phi^{-1}(1-\varepsilon) + \theta\right) \sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}} - b_1 \le 0 \right\},$$

the lower-level ALSO-X$\underline{\#}$ (6b) is equivalent to

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}, \boldsymbol{c}^\top \boldsymbol{x} \le t} \left\{ \boldsymbol{\mu}^\top \boldsymbol{x} + \left[\overline{G}\left(\left(\Phi^{-1}(1-\varepsilon)\right)^2 /2 \right) /\varepsilon + \theta \right] \sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}} - b_1 \right\},$$

where $\overline{G}(\tau) = G(\infty) - G(\tau)$ and $G(\tau) = \bar{k}\int_0^\tau \widehat{g}(z)dz$.

Then we study the exactness of ALSO-X$\underline{\#}$ for the following two conditions.

Condition I. For a single DRCCP under an elliptical reference distribution, suppose that $\boldsymbol{\Sigma} = \boldsymbol{\mu}\boldsymbol{\mu}^\top$ and $\boldsymbol{\mu}^\top \boldsymbol{x} \ge 0$ for any $\boldsymbol{x} \in \mathcal{X}$. In this case, we can simplify DRCCP (2) and the lower-level ALSO-X$\underline{\#}$ (6b) as

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \widehat{G}_\theta(\boldsymbol{\mu}^\top \boldsymbol{x}) = \left(1 + \Phi^{-1}(1-\varepsilon) + \theta\right) \boldsymbol{\mu}^\top \boldsymbol{x} - b_1 \le 0 \right\}, \tag{13a}$$

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} \left\{\overline{F}_\theta(\boldsymbol{\mu}^\top \boldsymbol{x}) = \left(1+\overline{G}\left(\left(\Phi^{-1}(1-\varepsilon)\right)^2/2\right)/\varepsilon+\theta\right)\boldsymbol{\mu}^\top\boldsymbol{x} - b_1 \colon \boldsymbol{c}^\top\boldsymbol{x} \le t\right\}, \tag{13b}$$

respectively. The exactness result readily follows from Theorem 8, which is summarized below.

Corollary 4 *Suppose that in a single DRCCP* (2)*, the reference distribution* $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ *is elliptical with affine mappings* $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$*,* $b_1(\boldsymbol{x}) = b_1$*,* $\boldsymbol{\Sigma} = \boldsymbol{\mu}\boldsymbol{\mu}^\top$*,* $\boldsymbol{\mu}^\top\boldsymbol{x} \ge 0$ *for any* $\boldsymbol{x} \in \mathcal{X}$*, and* $1+\Phi^{-1}(1-\varepsilon)+\theta \ge 0$*. Then* ALSO-X# *is exact.*

Proof According to the reformulations (13a) and (13b) and the assumptions that $\boldsymbol{\mu}^\top\boldsymbol{x} \ge 0$ for any $\boldsymbol{x} \in \mathcal{X}$ and $1+\Phi^{-1}(1-\varepsilon)+\theta \ge 0$, both functions $\widehat{G}_\theta(\cdot)$ and $\overline{F}_\theta(\cdot)$ are monotone nondecreasing. Hence, conditions in Theorem 8 are satisfied, and we have that ALSO-X# is exact. ☐

Condition II. For a single DRCCP under an elliptical reference distribution, suppose that $\mathcal{X} \subseteq \{0,1\}^n$, $\boldsymbol{\mu} \ge \mathbf{0}$, and $\boldsymbol{\Sigma} = \mathrm{Diag}(\boldsymbol{\mu})$. In this case, DRCCP (2) and the lower-level ALSO-X$\underline{\#}$ (6b) can be simplified as

$$v^* = \min_{\boldsymbol{x}\in\{0,1\}^n} \left\{\boldsymbol{c}^\top\boldsymbol{x} \colon \widehat{G}_\theta(\boldsymbol{\mu}^\top\boldsymbol{x}) = \boldsymbol{\mu}^\top\boldsymbol{x} + \left(\Phi^{-1}(1-\varepsilon)+\theta\right)\sqrt{\boldsymbol{\mu}^\top\boldsymbol{x}} - b_1 \le 0\right\}, \tag{14a}$$

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\{0,1\}^n} \left\{\overline{F}_\theta(\boldsymbol{\mu}^\top\boldsymbol{x}) = \boldsymbol{\mu}^\top\boldsymbol{x} + \left[\overline{G}\left(\left(\Phi^{-1}(1-\varepsilon)\right)^2/2\right)/\varepsilon+\theta\right]\sqrt{\boldsymbol{\mu}^\top\boldsymbol{x}} - b_1 \colon \boldsymbol{c}^\top\boldsymbol{x} \le t\right\}, \tag{14b}$$

respectively. According to Theorem 8, we have the following exactness result.

Corollary 5 *Suppose that in a single DRCCP* (2)*, the reference distribution* $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ *is elliptical with affine mappings* $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$*,* $b_1(\boldsymbol{x}) = b_1$*,* $\mathcal{X} \subseteq \{0,1\}^n$*,* $\boldsymbol{\mu} \ge \mathbf{0}$*,* $\boldsymbol{\Sigma} = \mathrm{Diag}(\boldsymbol{\mu})$*, and* $\Phi^{-1}(1-\varepsilon)+\theta \ge 0$*. Then* ALSO-X# *is exact.*

Proof According to the reformulations (14a) and (14b) and the assumptions that $\boldsymbol{\mu} \ge \mathbf{0}$ and $\Phi^{-1}(1-\varepsilon)+\theta \ge 0$, both functions $\widehat{G}_\theta(\cdot)$ and $\overline{F}_\theta(\cdot)$ are monotone nondecreasing. Hence, according to Theorem 8, we have that ALSO-X# is exact. ☐

We remark that the results in Corollary 4 and Corollary 5 hold for general type $q-$Wasserstein ambiguity set, as shown in Section 6.2.

5.2 Special Case II: Single DRCCPs with i.i.d. Random Parameters and Binary Decision

In this subsection, we study the exactness of a single DRCCP (2) with the binary decision variables, where we can provide the closed-form expression of the lower-level ALSO-X$\underline{\#}$.

Let us first consider a single packing DRCCP (2), where the deterministic set $\mathcal{X} \subseteq \{0,1\}^n$ is binary, the affine mappings are $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$, $b_1(\boldsymbol{x}) = b_1 \ge 0$, and support is nonnegative $\varXi \subseteq \mathbb{R}^n_+$ with i.i.d. random parameters $\tilde{\boldsymbol{\zeta}}$. That is, we consider the following DRCCP (2):

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{\boldsymbol{c}^\top\boldsymbol{x} \colon \inf_{\mathbb{P}\in\mathcal{P}_\infty} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} \colon \boldsymbol{x}^\top\tilde{\boldsymbol{\xi}} \le b_1\right\} \ge 1-\varepsilon\right\}.$$

In this case, DRCCP (2) is equivalent to

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{\boldsymbol{c}^\top\boldsymbol{x} \colon \widehat{G}_\theta(\boldsymbol{e}^\top\boldsymbol{x}) = 1-\varepsilon-\mathbb{P}\left\{\tilde{\boldsymbol{\zeta}} \colon \max_{\boldsymbol{\xi}\in\mathbb{R}^n_+}\left\{\sum_{i\in[n]}\xi_i x_i \colon \|\boldsymbol{\xi}-\tilde{\boldsymbol{\zeta}}\|_p \le \theta\right\} \le b_1\right\} \le 0\right\}. \tag{15a}$$

And the lower-level ALSO-X$\underline{\#}$ (6b) is equivalent to

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{c}^\top\boldsymbol{x}\le t} \left\{\overline{F}_\theta(\boldsymbol{e}^\top\boldsymbol{x}) = \mathrm{CVaR}_{1-\varepsilon}\left\{-b_1 + \max_{\boldsymbol{\xi}\in\mathbb{R}^n_+}\left[\sum_{i\in[n]}\xi_i x_i \colon \|\boldsymbol{\xi}-\tilde{\boldsymbol{\zeta}}\|_p \le \theta\right]\right\}\right\}. \tag{15b}$$

In this case, we can show that ALSO-X# is exact.

Corollary 6 *Consider a single DRCCP* (2) *with affine mappings* $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$, $b_1(\boldsymbol{x}) = b_1 \geq 0$, *the deterministic set* $\mathcal{X} \subseteq \{0,1\}^n$, *and the random parameters* $\tilde{\boldsymbol{\zeta}}$ *are i.i.d. and nonnegative. Then* ALSO-X# *is exact.*

Proof It is sufficient to show that functions $\widehat{G}_\theta(\cdot)$ and $\overline{F}_\theta(\cdot)$ indeed exist and share the same monotonicity. We first notice that

$$\max_{\boldsymbol{\xi} \in \mathbb{R}^n_+} \left\{ \sum_{i\in[n]} \xi_i x_i \colon \|\boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}}\|_p \leq \theta \right\} = \sum_{i\in[n]} \tilde{\zeta}_i x_i + \max_{\boldsymbol{\xi} \in \mathbb{R}^n_+} \left\{ \sum_{i\in[n]} \xi_i x_i \colon \|\boldsymbol{\xi}\|_p \leq \theta \right\}.$$

Since $\{\tilde{\zeta}_i\}_{i\in[n]}$ are i.i.d. nonnegative random parameters, for any $\boldsymbol{x} \in \mathcal{X}$ such that $\boldsymbol{e}^\top \boldsymbol{x} = \ell$, we have

$$\max_{\boldsymbol{\xi} \in \mathbb{R}^n_+} \left\{ \sum_{i\in[n]} \xi_i x_i \colon \|\boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}}\|_p \leq \theta \right\} \overset{\mathbb{P}_{\tilde{\xi}}}{\sim} \sum_{i\in[\ell]} \tilde{\zeta}_i + \max_{\boldsymbol{\xi} \in \mathbb{R}^\ell_+} \left\{ \sum_{i\in[\ell]} \xi_i \colon \|\boldsymbol{\xi}\|_p \leq \theta \right\}.$$

Hence, let us define

$$\widehat{G}_\theta(\boldsymbol{e}^\top \boldsymbol{x}) = \widehat{G}_\theta(\ell) = 1 - \varepsilon - \mathbb{P}\left\{ \tilde{\boldsymbol{\zeta}} \colon \max_{\boldsymbol{\xi} \in \mathbb{R}^n_+} \left\{ \sum_{i\in[n]} \xi_i x_i \colon \|\boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}}\|_p \leq \theta \right\} \leq b_1 \right\}.$$

Similarly, we also define

$$\overline{F}_\theta(\boldsymbol{e}^\top \boldsymbol{x}) = \overline{F}_\theta(\ell) = \mathrm{CVaR}_{1-\varepsilon} \left\{ -b_1 + \max_{\boldsymbol{\xi} \in \mathbb{R}^n_+} \left[\sum_{i\in[n]} \xi_i x_i \colon \|\boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}}\|_p \leq \theta \right] \right\}.$$

Since all the random parameters $\{\tilde{\zeta}_i\}_{i\in[n]}$ are nonnegative, we have

$$\sum_{i\in[\ell+1]} \tilde{\zeta}_i + \max_{\boldsymbol{\xi} \in \mathbb{R}^{\ell+1}_+} \left\{ \sum_{i\in[\ell+1]} \xi_i \colon \|\boldsymbol{\xi}\|_p \leq \theta \right\} \geq \sum_{i\in[\ell]} \tilde{\zeta}_i + \max_{\boldsymbol{\xi} \in \mathbb{R}^\ell_+} \left\{ \sum_{i\in[\ell]} \xi_i \colon \|\boldsymbol{\xi}\|_p \leq \theta \right\}$$

almost surely. Thus, we have

$$\widehat{G}_\theta(\ell) \leq \widehat{G}_\theta(\ell+1), \overline{F}_\theta(\ell) \leq \overline{F}_\theta(\ell+1).$$

According to Theorem 8, ALSO-X# is exact. □

We remark that the result in Corollary 6 also holds for a covering DRCCP. That is, let us consider the following

$$v^* = \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \inf_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} \colon \boldsymbol{x}^\top \tilde{\boldsymbol{\xi}} \geq b_1 \right\} \geq 1 - \varepsilon \right\},$$

where the random parameters $\{\boldsymbol{\zeta}_i\}_{i\in[n]}$ are i.i.d. and nonnegative and cost vector $\boldsymbol{c}$ is nonnegative. Let us denote $1 - y_i = x_i$ for all $i \in [n]$. Then chance constrained covering problem is equivalent to

$$v^* = \min_{(\boldsymbol{e}-\boldsymbol{y}) \in \mathcal{X}} \left\{ \boldsymbol{c}^\top (\boldsymbol{e} - \boldsymbol{y}) \colon \inf_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} \colon (\boldsymbol{e} - \boldsymbol{y})^\top \tilde{\boldsymbol{\xi}} \geq b_1 \right\} \geq 1 - \varepsilon \right\}.$$

As a result, the proof in Corollary 6 simply follows, which is summarized below.

Corollary 7 *Consider a single DRCCP with affine mappings* $\boldsymbol{a}_1(\boldsymbol{x}) = -\boldsymbol{x}$, $b_1(\boldsymbol{x}) = -b_1 \leq 0$, *the deterministic set* $\mathcal{X} \subseteq \{0,1\}^n$, *and the random parameters* $\tilde{\boldsymbol{\zeta}}$ *being i.i.d. and nonnegative. Then* ALSO-X# *is exact.*

5.3 Special Case III: Single DRCCPs with Empirical Reference Distribution

In this subsection, we study the exactness of a single DRCCP with empirical reference distribution, where the affine mappings are $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$, $b_1(\boldsymbol{x}) = b_1 \geq 0$, and the support is discrete. That is, we consider the following DRCCP:

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \mathbb{P}\left\{ \tilde{\boldsymbol{\zeta}} \colon \theta \left\|\boldsymbol{x}\right\|_* + \boldsymbol{x}^\top \tilde{\boldsymbol{\zeta}} \leq b_1 \right\} \geq 1-\varepsilon \right\}. \tag{16}$$

Corollary 8 *Consider a single DRCCP* (2) *with affine mappings* $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$, $b_1(\boldsymbol{x}) = b_1 \geq 0$, *and the norm* $\|\cdot\|$ *is the generalized Mahalanobis norm associated with the matrix* $\boldsymbol{\Sigma} = \boldsymbol{\mu}\boldsymbol{\mu}^\top$ *and* $\boldsymbol{\mu}^\top \boldsymbol{x} \geq 0$ *for any* $\boldsymbol{x} \in \mathcal{X}$. *Suppose the support is discrete with* $\mathbb{P}\{\tilde{\boldsymbol{\zeta}} = \zeta_i \boldsymbol{\mu}\} = p_i \geq 0$ *for all* $i \in [N]$, *where* $\sum_{i\in[N]} p_i = 1$ *and* $0 \leq \zeta_1 < \zeta_2 < \cdots < \zeta_N$ *are scalars. Then* ALSO-X# *is exact.*

Proof We define $K \in [N]$ as $\sum_{i\in[K-1]} p_i < 1-\varepsilon, \sum_{i\in[K]} p_i \geq 1-\varepsilon$. By definition, we have $\zeta_K > 0$ and $\|\boldsymbol{x}\|_* = |\boldsymbol{\mu}^\top \boldsymbol{x}|$. According to the assumption that $\mathbb{P}\{\tilde{\boldsymbol{\zeta}} = \zeta_i \boldsymbol{\mu}\} = p_i \geq 0$ for all $i \in [N]$, DRCCP (16) can be simplified as

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \widehat{G}_\theta(\boldsymbol{\mu}^\top \boldsymbol{x}) = 1-\varepsilon - \sum_{i\in[N]} p_i \mathbb{I}\left((\zeta_i + \theta)\, \boldsymbol{\mu}^\top \boldsymbol{x} \leq b_1 \right) \leq 0 \right\},$$

and the corresponding lower-level ALSO-X$\underline{\#}$ (6b) is equivalent to

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}, \boldsymbol{c}^\top\boldsymbol{x}\leq t} \left\{ \overline{F}_\theta(\boldsymbol{\mu}^\top \boldsymbol{x}) = \frac{1}{\varepsilon} \left[\left(\sum_{i\in[K]} p_i - (1-\varepsilon) \right) \zeta_i + \sum_{j\in[K+1,N]} p_j \zeta_j \right] \boldsymbol{\mu}^\top \boldsymbol{x} + \theta \|\boldsymbol{\mu}\|_2^{-1} \boldsymbol{\mu}^\top \boldsymbol{x} - b_1 \right\}.$$

Both functions $\widehat{G}_\theta(\cdot)$ and $\overline{F}_\theta(\cdot)$ are monotone nondecreasing. Therefore, the conditions in Theorem 8 are satisfied, and we conclude that ALSO-X# is exact. □

6 Extensions: DRCCPs under Type $q-$Wasserstein Ambiguity Set

In this section, we extend our discussions to type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$. We first provide equivalent reformulations. Then we show that under type $q-$Wasserstein ambiguity set, ALSO-X# can provide an optimal solution to a DRCCP. The results in this section rely on the equivalent reformulations of DRCCP, ALSO-X, CVaR approximation, and ALSO-X# under type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$, which are displayed in Appendix B.

6.1 Comparisons of ALSO-X#, ALSO-X$\underline{\#}$, ALSO-X, and CVaR Approximation

As mentioned in Section 3, the main result of this paper in Theorem 1 cannot be extended to type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$. That is, ALSO-X and ALSO-X# are not comparable under type $q-$Wasserstein ambiguity set when $q \in [1, \infty)$. Below is an example.

Example 6 Consider a single DRCCP under type $1-$Wasserstein ambiguity set with $\theta = 1$ and $\|\cdot\|_* = \|\cdot\|_2$. Assume that the empirical distribution has 4 equiprobable scenarios (i.e., $N = 4$, $\mathbb{P}\{\tilde{\boldsymbol{\zeta}} = \boldsymbol{\zeta}^i\} = 1/N$), risk parameter $\varepsilon = 1/2$, the deterministic set $\mathcal{X} = \mathbb{R}^3_+$, $\boldsymbol{c} = (-4,-2,-3)^\top$, function $\boldsymbol{a}_1(\boldsymbol{x})^\top \boldsymbol{\zeta} - b_1(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\zeta} - 3$, $\boldsymbol{\zeta}^1 = (4,6,3)^\top$, $\boldsymbol{\zeta}^2 = (5,0,3)^\top$, $\boldsymbol{\zeta}^3 = (2,1,4)^\top$, and $\boldsymbol{\zeta}^4 = (0,2,5)^\top$. In this example, numerically, we can solve ALSO-X v_1^A, ALSO-X# $v_1^{A\#}$, and CVaR approximation v_1^{CVaR}, where the approximated objective values are $v_1^A = -2.4929$, $v_1^{A\#} = -2.4369$, and $v_1^{\text{CVaR}} = -2.033$ with error bound $[-10^{-4}, 10^{-4}]$. ◇

Albeit ALSO-X and ALSO-X# are not comparable, following the similar proofs as those of Theorem 2 and Theorem 3 in Section 3, we can prove that ALSO-X# is better than CVaR approximation and ALSO-X# is better than ALSO-X$\underline{\#}$ under type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$. Interested readers are referred to Appendix $\overline{\text{B}}$ for proofs.

Proposition 1 *Under type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$, suppose that for any objective upper bound t such that $t \geq \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$,* ALSO-X# *is better than* CVaR *approximation.*

Proposition 2 *Under type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$, suppose that for any objective upper bound t such that $t \geq \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and the lower-level* ALSO-X$\underline{\#}$ *admits a unique optimal $\boldsymbol{x}$-solution,* ALSO-X# *is better than* ALSO-X$\underline{\#}$.

6.2 Exactness of ALSO-X#

In this subsection, we extend the discussion in Section 5.1 for the single DRCCP (22) under elliptical distribution with affine mappings $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$ and $b_1(\boldsymbol{x}) = b_1$. Due to the page limit, we refer interested readers to Appendix B for detailed derivations. Under type $q-$Wasserstein ambiguity set with elliptical reference distribution, DRCCP resorts to

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \boldsymbol{\mu}^\top \boldsymbol{x} + \eta_q^* \sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}} - b_1 \leq 0 \right\}, \tag{17}$$

with

$$\eta_q^* = \min_{\eta} \left\{ \eta \colon \int_{\Phi^{-1}(1-\varepsilon)}^{\eta} (\eta - t)^q \, \bar{k}\widehat{g}(t^2/2) dt \geq \theta^q, \eta \geq \Phi^{-1}(1-\varepsilon) \right\}.$$

In this case, the lower-level ALSO-X$\underline{\#}$ is equivalent to

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}, \boldsymbol{c}^\top \boldsymbol{x} \leq t} \left\{ \boldsymbol{\mu}^\top \boldsymbol{x} + \left[\overline{G} \left(\left(\Phi^{-1}(1-\varepsilon) \right)^2 /2 \right) /\varepsilon + \theta \varepsilon^{-\frac{1}{q}} \right] \sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}} - b_1 \right\}. \tag{18}$$

Let us make the same assumption as that in Condition I of Section 5.1, i.e., for a single DRCCP under an elliptical reference distribution, suppose that $\boldsymbol{\Sigma} = \boldsymbol{\mu}\boldsymbol{\mu}^\top$ and $\boldsymbol{\mu}^\top \boldsymbol{x} \geq 0$ for any $\boldsymbol{x} \in \mathcal{X}$. Then DRCCP (17) and the lower-level ALSO-X$\underline{\#}$ (18) can be simplified as

$$v^* = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \widehat{G}_\theta(\boldsymbol{\mu}^\top \boldsymbol{x}) = \left(1 + \eta_q^*\right) \boldsymbol{\mu}^\top \boldsymbol{x} - b_1 \leq 0 \right\},$$
$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} \left\{ \overline{F}_\theta(\boldsymbol{\mu}^\top \boldsymbol{x}) = \left(1 + \overline{G} \left(\left(\Phi^{-1}(1-\varepsilon) \right)^2 /2 \right) /\varepsilon + \theta \varepsilon^{-\frac{1}{q}} \right) \boldsymbol{\mu}^\top \boldsymbol{x} - b_1 \colon \boldsymbol{c}^\top \boldsymbol{x} \leq t \right\},$$

respectively. The assumptions that $\boldsymbol{\mu}^\top \boldsymbol{x} \geq 0$ for any $\boldsymbol{x} \in \mathcal{X}$ and $1 + \eta_q^* \geq 0$ ensure that both functions $\widehat{G}_\theta(\cdot)$ and $\overline{F}_\theta(\cdot)$ are monotone nondecreasing. Then the exactness result directly follows from Theorem 8 and Corollary 4, which is summarized below.

Corollary 9 *Suppose that in a single DRCCP* (17)*, the reference distribution $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ is elliptical with affine mappings $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$, $b_1(\boldsymbol{x}) = b_1$, $\boldsymbol{\Sigma} = \boldsymbol{\mu}\boldsymbol{\mu}^\top$, $\boldsymbol{\mu}^\top \boldsymbol{x} \geq 0$ for any $\boldsymbol{x} \in \mathcal{X}$, and $\eta_q^* \geq -1$. Then* ALSO-X# *is exact.*

Similarly, let us make the same assumption as that in Condition II of Section 5.1, i.e., for a single DRCCP under an elliptical reference distribution, suppose that $\mathcal{X} \subseteq \{0,1\}^n$, $\boldsymbol{\mu} \geq \boldsymbol{0}$, and $\boldsymbol{\Sigma} = \operatorname{Diag}(\boldsymbol{\mu})$. In this case, DRCCP (17) and the lower-level ALSO-X$\underline{\#}$ (18) can be simplified as

$$v^* = \min_{\boldsymbol{x}\in\{0,1\}^n} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \widehat{G}_\theta(\boldsymbol{\mu}^\top \boldsymbol{x}) = \boldsymbol{\mu}^\top \boldsymbol{x} + \eta_q^* \sqrt{\boldsymbol{\mu}^\top \boldsymbol{x}} - b_1 \leq 0 \right\},$$

$$\boldsymbol{x}^* \in \underset{\boldsymbol{x}\in\{0,1\}^n}{\operatorname{argmin}} \left\{ \overline{F}_\theta(\boldsymbol{\mu}^\top \boldsymbol{x}) = \boldsymbol{\mu}^\top \boldsymbol{x} + \left[\overline{G}\left(\left(\Phi^{-1}(1-\varepsilon)\right)^2/2 \right)/\varepsilon + \theta\varepsilon^{-\frac{1}{q}} \right] \sqrt{\boldsymbol{\mu}^\top \boldsymbol{x}} - b_1 \colon \boldsymbol{c}^\top \boldsymbol{x} \le t \right\}.$$

The assumptions $\boldsymbol{\mu} \ge \mathbf{0}$ and $\eta_q^* \ge 0$ guarantee that both functions $\widehat{G}_\theta(\cdot)$ and $\overline{F}_\theta(\cdot)$ are monotone nondecreasing. Then we have the following exactness result.

Corollary 10 *Suppose that in a single DRCCP* (17)*, the reference distribution* $\mathbb{P}_{\tilde{\zeta}}$ *is elliptical with affine mappings* $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$*,* $b_1(\boldsymbol{x}) = b_1$*,* $\mathcal{X} \subseteq \{0,1\}^n$*,* $\boldsymbol{\mu} \ge \mathbf{0}$*,* $\boldsymbol{\Sigma} = \mathrm{Diag}(\boldsymbol{\mu})$*, and* $\eta_q^* \ge 0$*. Then* ALSO-X# *is exact.*

7 Numerical Study

In this section, we numerically demonstrate the efficacy of the proposed methods. All the instances in this section are executed in Python 3.9 with calls to solver Gurobi (version 9.1.1 with default settings) on a personal PC with an Apple M1 Pro processor and 16G of memory.

7.1 Synthetic Cases

We evaluate the differences among CVaR approximation, ALSO-X, and ALSO-X# using "Improvement from CVaR approximation" to denote the percentage of differences between the value of a proposed algorithm and CVaR approximation, i.e.,

$$\begin{aligned}&\text{Improvement from CVaR approximation } (\%)\\&= \frac{\text{CVaR approximation value} - \text{output value of a proposed algorithm}}{|\text{CVaR approximation value}|} \times 100.\end{aligned}$$

Typically, CVaR approximation is quite conservative. As a better alternative, we also use "Improvement from ALSO-X" to denote the percentage of differences between the value of a proposed algorithm and ALSO-X approximation, i.e.,

$$\text{Improvement from ALSO-X } (\%) = \frac{\text{ALSO-X value} - \text{output value of a proposed algorithm}}{|\text{ALSO-X value}|} \times 100.$$

We compare the performances of CVaR approximation, ALSO-X, and ALSO-X# of solving a single DRCCP with different sizes of data points $N = 400, 600, 1000$, varying risk level $\varepsilon = 0.10, 0.20$, fixed Wasserstein radius $\theta = 0.05$, and also different dimensions of decision variables $n = 20, 40, 100$. In ALSO-X and ALSO-X# algorithm, we use the optimal value from CVaR approximation as an initial upper bound t_U and the quantile bound from [1] as an initial lower bound t_L. For each parametric setting, we generate 5 random instances and report their average performance.

We separate our discussions into type $\infty-$Wasserstein ambiguity set and type $2-$Wasserstein ambiguity set, respectively.

Case I. Testing a DRCCP with type $\infty-$Wasserstein ambiguity set. We split the discussions into the continuous case and the binary case.

1.1 Continuous Case. Let us first consider the following DRCCP:

$$v^* = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \boldsymbol{x} \in [0,1]^n, \frac{1}{N} \sum_{i\in[N]} \mathbb{1} \left[\theta \|\boldsymbol{x}\|_2 + \sum_{j\in[n]} \zeta_j^i x_j \le b^i \right] \ge 1 - \varepsilon \right\}.$$

Above, we generate the samples $\{\boldsymbol{\zeta}^i\}_{i\in[N]}$ by assuming that the random parameters $\tilde{\boldsymbol{\zeta}}$ are discrete and i.i.d. uniformly distributed between 1 and 80. We set $\delta_1 = 10^{-2}$ in ALSO-X# Algorithm 1. For each random instance, we assume the cost vector $\boldsymbol{c}$ to be random integer with each entry uniformly distributed between -30 and -1. And we assume the random parameter $\tilde{b}$ is discrete and i.i.d. uniformly distributed between 1 and 20. The numerical results are displayed in Table 1. We see that although the computation time of ALSO-X# is comparable to that of ALSO-X, the solution quality of ALSO-X# is around 4%-10% better than that of ALSO-X. This demonstrates the effectiveness of our proposed ALSO-X#.

Table 1: Numerical Results of a DRCCP under Type $\infty-$Wasserstein Ambiguity Set with $\theta = 0.05$

ε	N	n	CVaR	ALSO-X		ALSO-X#	
			Time (s)	Time (s)	Improvement from CVaR approximation (%)	Time (s)	Improvement from ALSO-X (%)
0.10	400	20	0.15	1.53	13.28	1.56	8.73
		40	0.29	3.26	11.29	2.73	5.86
		100	0.73	9.12	14.59	13.62	6.77
	600	20	0.39	2.27	8.48	2.78	6.06
		40	0.35	4.64	9.22	4.41	6.38
		100	0.78	11.82	9.79	12.73	5.69
	1000	20	0.25	4.77	10.76	4.81	9.16
		40	0.72	7.20	9.16	7.35	5.10
		100	1.62	20.15	6.31	21.98	3.38
0.20	400	20	0.22	2.75	12.94	2.43	9.20
		40	0.41	2.74	13.53	2.34	6.28
		100	0.66	4.79	5.39	5.51	6.02
	600	20	0.49	3.49	6.49	3.24	7.91
		40	0.54	6.18	5.33	4.07	5.97
		100	1.43	12.87	3.82	10.66	7.76
	1000	20	0.52	6.50	8.59	5.26	5.76
		40	0.84	7.68	4.21	7.53	5.09
		100	2.38	21.92	4.76	18.47	3.73

1.2 Binary Case. Let us first consider the following DRCCP:

$$v^* = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \boldsymbol{x} \in \{0,1\}^n, \frac{1}{N} \sum_{i \in [N]} \mathbb{1} \left[\theta \left\| \boldsymbol{x} \right\|_1 + \sum_{j \in [n]} \zeta_j^i x_j \le b^i \right] \ge 1 - \varepsilon \right\}.$$

Above, we generate the samples $\{\boldsymbol{\zeta}^i\}_{i\in[N]}$ by assuming that the random parameters $\tilde{\boldsymbol{\zeta}}$ are discrete and i.i.d. uniformly distributed between -10 and 20. We set $\delta_1 = 0.5$ in ALSO-X# Algorithm 1. For each random instance, we assume the cost vector $\boldsymbol{c}$ to be random integer, with each entry uniformly distributed between -10 and -1. And we assume the random parameter $\tilde{b}$ is discrete and i.i.d. uniformly distributed between 1 and 200. The numerical results for this case are displayed in Table 2. Similar to the results in Table 1, we conclude that ALSO-X# enhances the solution quality from ALSO-X by around 3%-7% improvement with a comparable computation time.

Case II. Testing a DRCCP with type $2-$Wasserstein ambiguity set. We split the discussions into the continuous case and the binary case.

2.1 Continuous Case. Let us first consider the following DRCCP:

$$v^* = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \boldsymbol{x} \in [0,1]^n, \inf_{\mathbb{P} \in \mathcal{P}_2} \mathbb{P} \left\{ \tilde{\boldsymbol{\xi}} \colon \boldsymbol{x}^\top \tilde{\boldsymbol{\xi}} \le b \right\} \ge 1 - \varepsilon \right\}.$$

Note that this DRCCP may not have an MIP reformulation (see, e.g., [27]). Above, we generate the samples $\{\boldsymbol{\zeta}^i\}_{i\in[N]}$ by assuming that the random parameters $\tilde{\boldsymbol{\zeta}}$ are discrete and i.i.d. uniformly distributed between 1 and 80. We set $\delta_1 = 10^{-2}$ in ALSO-X# Algorithm 1. For each random instance, we assume $b = 10$ and assume the cost vector $\boldsymbol{c}$ to be random integer with each entry uniformly distributed between -20 and -1. The numerical results are displayed in Table 3. We show that ALSO-X# yields around 5%-11% improvement. We also notice that ALSO-X can improve the solution of CVaR approximation.

2.2 Binary Case. Let us first consider the following DRCCP:

$$v^* = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \boldsymbol{x} \in \{0,1\}^n, \inf_{\mathbb{P} \in \mathcal{P}_2} \mathbb{P} \left\{ \tilde{\boldsymbol{\xi}} \colon \boldsymbol{x}^\top \tilde{\boldsymbol{\xi}} \le b \right\} \ge 1 - \varepsilon \right\}.$$

Table 2: Numerical Results of a Binary DRCCP under Type $\infty-$Wasserstein Ambiguity Set with $\theta = 0.05$

ε	N	n	CVaR	ALSO-X		ALSO-X#	
			Time (s)	Time (s)	Improvement from CVaR approximation (%)	Time (s)	Improvement from ALSO-X (%)
0.10	400	20	0.30	1.97	5.16	2.54	6.67
		40	0.31	5.65	8.03	8.34	7.10
		100	0.24	7.09	3.09	7.97	4.36
	600	20	0.31	5.28	7.49	4.75	5.33
		40	1.29	12.18	5.11	15.50	7.14
		100	1.35	14.35	5.19	16.23	5.56
	1000	20	0.95	5.82	3.51	6.29	2.78
		40	0.52	16.13	3.63	17.48	2.83
		100	0.61	24.69	4.82	26.11	3.19
0.20	400	20	0.40	5.75	4.73	5.59	7.35
		40	0.57	8.14	4.27	9.43	7.32
		100	0.64	9.34	4.04	11.48	5.41
	600	20	0.42	5.58	4.79	5.24	5.08
		40	0.80	16.92	3.37	14.95	6.92
		100	1.45	21.12	4.58	21.06	6.36
	1000	20	1.04	7.51	3.49	7.45	6.25
		40	1.26	24.39	4.12	22.99	6.24
		100	2.70	32.93	2.87	33.65	5.56

Table 3: Numerical Results of a DRCCP under Type $2-$Wasserstein Ambiguity Set with $\theta = 0.50$

ε	N	n	CVaR	ALSO-X		ALSO-X#	
			Time (s)	Time (s)	Improvement from CVaR approximation (%)	Time (s)	Improvement from CVaR approximation (%)
0.10	400	20	0.09	0.90	7.21	0.81	7.22
		40	0.25	2.42	6.43	2.06	6.54
		100	0.55	5.76	4.73	3.97	4.73
	600	20	0.20	1.40	8.60	1.12	8.60
		40	0.38	3.83	6.92	2.72	6.92
		100	1.16	12.46	5.27	8.96	5.27
	1000	20	0.23	2.24	8.78	1.88	8.82
		40	0.76	5.62	7.82	4.39	7.94
		100	1.48	18.82	5.82	13.36	5.87
0.20	400	20	0.25	0.93	10.45	0.82	10.61
		40	0.45	2.89	10.17	2.16	10.24
		100	0.66	6.17	6.14	4.03	6.32
	600	20	0.13	1.41	10.10	1.19	10.10
		40	0.32	3.93	8.46	2.83	8.62
		100	1.14	13.38	6.77	9.30	6.76
	1000	20	0.28	2.35	10.70	2.05	10.70
		40	0.56	5.67	10.33	4.29	10.33
		100	1.83	20.01	6.82	14.30	6.92

Above, we generate the samples $\{\boldsymbol{\zeta}^i\}_{i\in[N]}$ by assuming that the random parameters $\tilde{\boldsymbol{\zeta}}$ are discrete and i.i.d. uniformly distributed between -20 and 50. We set $\delta_1 = 0.5$ in ALSO-X# Algorithm 1. For each random instance, we assume $b = 400$ and assume the cost vector $\boldsymbol{c}$ to be random integer, with each entry uniformly distributed between -20 and -10. In this numerical experiment, we suppose the dimensions of decision variables $n = 20, 40$. The numerical results for this case are displayed in Table 4. We conclude that ALSO-X# enhances the solution quality from CVaR approximation by around 5-10% improvement.

7.2 Application: Resource Allocation in Wireless Communication

We compare ALSO-X# and the exact method in the wireless communication network problem, where we can use ALSO-X# to minimize the energy consumed. Specifically, we consider a predictive resource allocation problem for energy-efficient video streaming (see, e.g., [3, 4]), where chance constraints are employed to ensure a high quality of service for each user. The objective of this problem is to minimize the energy consumption in transmitting the video content to the users while satisfying the chance

Table 4: Numerical Results of a Binary DRCCP under Type 2−Wasserstein Ambiguity Set with $\theta = 0.20$

ε	N	n	CVaR	ALSO-X		ALSO-X#	
			Time (s)	Time (s)	Improvement from CVaR approximation (%)	Time (s)	Improvement from CVaR approximation (%)
0.10	400	20	0.23	1.55	5.80	1.43	6.02
		40	2.81	34.11	8.23	25.23	7.10
	600	20	0.36	2.60	5.19	2.69	5.19
		40	5.77	57.60	9.81	55.39	10.19
	1000	20	1.44	5.29	7.92	4.78	8.19
		40	12.45	86.86	9.67	85.98	9.67
0.20	400	20	0.24	1.80	7.67	1.65	7.67
		40	3.49	35.70	9.46	27.71	9.46
	600	20	0.57	3.14	7.73	3.08	7.82
		40	13.85	60.79	9.82	74.74	11.50
	1000	20	1.01	5.61	7.26	4.97	7.66
		40	16.48	82.65	8.73	87.72	9.87

constraints. The problem is formally formulated as

$$\min_{\boldsymbol{x}\in[0,1]^{n\times T}} \left\{ \sum_{t\in[T]}\sum_{i\in[n]} x_{i,t} \colon \inf_{\mathbb{P}\in\mathcal{P}_\infty} \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} \colon \sum_{t'\in[t]} \tilde{\xi}_{i,t'} x_{i,t'} \geq D_{i,t}, \forall t\in[T] \right\} \geq 1-\varepsilon, \forall i\in[n], \right.$$
$$\left. \sum_{i\in[n]} x_{i,t} \leq 1, \forall t\in[T] \right\}, \tag{21}$$

where $x_{i,t}$ denotes resource allocation decision at time slot t to user i, n denotes the number of all users, T denotes the number of time slots, and $D_{i,t}$ denotes the demand for each user $i \in [n]$ up to time $t \in [T]$. The random parameter $\xi_{i,t}$ denotes the random amount of available rate for user $i \in [n]$ at time slot t.

Above, we generate the samples $\{\boldsymbol{\zeta}^i\}_{i\in[N]}$ by assuming that the random parameters $\tilde{\boldsymbol{\zeta}}$ are discrete and i.i.d. uniformly distributed between 20 and 40. We set $\delta_1 = 10^{-1}$ in ALSO-X# Algorithm 1. And we assume for each user $i \in [n]$ up to time $t \in [T]$, the demand $D_{i,t}$ is $D_{i,t} = tD$ with $D = 1.0, 1.5$. We consider the number of data samples $N = 56, 72$, the risk level $\varepsilon = 0.20, 0.30$, the Wasserstein radius $\theta = 0.50$, the number of users $n = 8$, and the number of time slots $T = 60$. We generate 5 random instances to find their average performance.

The proposed ALSO-X# can effectively identify better feasible solutions than the exact Big-M model with a much shorter solution time, which is typically required in many wireless communication applications. Since we consider the number of time slots $T = 60$s, for a fair comparison, we set the time limit of each instance to 60s (i.e., 1 minute), and we use "UB" and "LB" to denote the best upper bound and the best lower bound found by the Big-M model within the one-minute time limit. Since we may not be able to solve the Big-M model to optimality in one minute, we use GAP to denote its optimality gap as GAP (%) $= (|\text{UB} - \text{LB}|)/(|\text{LB}|) \times 100$, and we use the term "Improvement from Big-M model" to denote the solution quality improvement of ALSO-X#, i.e., Improvement from Big-M model (%) $=$ (UB $-$ ALSO-X# value)$/(|\text{UB}|) \times 100$. The numerical results for this case are displayed in Table 5. We find that ALSO-X# can provide better solutions than the Big-M model in a much shorter time, which validates the efficacy of our proposed methods. We remark that we can design real-time algorithms by using ALSO-X# as a future study (see, e.g., [4–6]).

Subsequently, we demonstrate how to use the ALSO-X# result to save energy-consuming. For illustration, we consider the following parametric setting $D = 1.0$, $N = 56$, $\varepsilon = 0.20$, and $n = 8$. We first numerically choose a proper Wasserstein radius. We use generated data to solve the DRCCP with ALSO-X#, the DRCCP with Big-M model, and the regular CCP counterpart (i.e., $\theta = 0$) with ALSO-X#, respectively. Then we generate new samples with the same sample size and obtain 95% confidence intervals by plugging the solutions in the regular CCP to calculate the corresponding probability that the constraints are satisfied. We repeat the same procedure for a list of θ values and select the smallest θ that the confidence interval of the violation probability in the DRCCP is beyond that of the regular CCP.

Table 5: Comparisons of ALSO-X# and Big-M Model in Resource Allocation Problem (21)

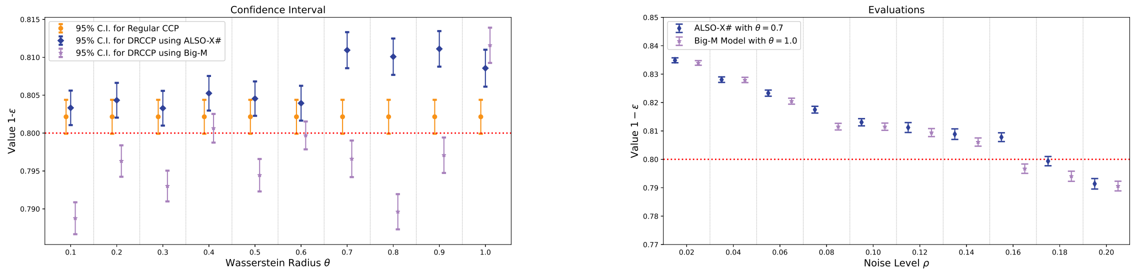
D	ε	N	Big-M Model		ALSO-X#	
			Gap (%)	Time (s)	Improvement from Big-M model (%)	Time (s)
1.0	0.20	56	19.05	60	0.10	21.41
		72	19.88	60	0.12	36.48
	0.30	56	25.53	60	0.10	25.35
		72	28.13	60	0.21	37.07
1.5	0.20	56	19.60	60	0.14	34.19
		72	20.68	60	0.19	41.42
	0.30	56	27.50	60	0.23	36.93
		72	28.84	60	0.39	42.06

Specifically, to select the smallest θ, we take the following steps: (i) for each $\theta \in \{0.1, 0.2, \cdots, 0.9, 1.0\}$, we generate $N = 56$ scenarios and solve the DRCCP and its regular CCP counterpart; (ii) generate $N = 56$ new scenarios with the same parameters; (iii) plug the solution from part (i) into the newly generated scenarios and calculate the probability that the constraints are satisfied; (iv) repeat previous procedures 50 times and output the 95% confidence intervals; and (v) choose the smallest θ such that the confidence interval of the DRCCP is entirely above that of the regular CCP counterpart. The result is shown in Figure 2a. The DRCCP using ALSO-X# with Wasserstein radius $\theta = 0.7$ can guarantee that the chance constraints are satisfied with a high probability. In contrast, the best Wasserstein radius is $\theta = 1.0$ when applying the result from the Big-M model. In this way, we conclude that compared with the Big-M model, the average energy saving using the ALSO-X# is 1.93%.

With the best-tuned Wasserstein radius, we compare the performances of the solutions from the Big-M model and ALSO-X# by generating 100 new scenarios to evaluate the probability that the constraints are satisfied and output the corresponding 95% confidence intervals. The testing setting is the same as that of the training setting above, except that we assume that random parameters $\tilde{\boldsymbol{\zeta}}$ are discrete and i.i.d. uniformly distributed between $20 \times (1 - 1.8\rho)$ and $40 \times (1 + 0.4\rho)$ with $\rho \in [0.02, 0.20]$, where the value ρ represents the noise level in the training distribution. The result is displayed in Figure 2b. It is seen that compared with the Big-M model, ALSO-X# yields ideal lower constraint violations. Specifically, when the noise level is small, i.e., $0.02 \leq \rho \leq 0.16$ in Figure 2b, ALSO-X# often guarantees a lower violation of constraints. However, when the noise level is large, i.e., $\rho = 0.20$ in Figure 2b, both ALSO-X# and Big-M model cannot provide the chance constraint guarantee. This result further demonstrates that using ALSO-X# with the best-tune Wasserstein radius can be better or at least achieve the similar constraint violation probability as the Big-M model. As mentioned in the previous paragraph, ALSO-X# provides a better solution with lower energy consumption. This demonstrates the better solution quality of ALSO-X# compared to the Big-M model.

8 Conclusion

In this work, we proposed ALSO-X# for solving distributionally robust chance constrained programs (DRCCPs) by integrating ALSO-X and CVaR approximation. We proved that ALSO-X# is better than CVaR approximation even when this deterministic set is nonconvex. We provided sufficient conditions that ALSO-X# always outperforms ALSO-X, i.e., when ALSO-X admits a unique optimal solution. We showed that ALSO-X# can deliver an optimal solution to a DRCCP and also extended the discussions to the general Wasserstein ambiguity set. Our numerical studies demonstrated the effectiveness of ALSO-X#. For a future study, it will be interesting to implement the ALSO-X# algorithm in real-time for wireless communication problems. Another interesting direction is to study the approximation guarantees of ALSO-X# when solving a DRCCP.



(a) Comparisons of Tuning Wasserstein Radii

(b) Evaluations of ALSO-X# and Big-M Model with the Best-tuned Wasserstein Radii

Fig. 2: Tuning Wasserstein Radius and Comparing the ALSO-X# and Big-M Model Solutions in Resource Allocation (21)

Acknowledgment

This research has been supported in part by the National Science Foundation grants 2246414 and 2246417.

Declarations

Funding and/or Conflicts of interests/Competing interests: This research has been supported by the National Science Foundation. Authors have no other competing interests to report.

References

1. S. Ahmed, J. Luedtke, Y. Song, and W. Xie. Nonanticipative duality, relaxations, and formulations for chance-constrained stochastic programs. *Mathematical Programming*, 162(1):51–81, 2017.
2. S. Ahmed and A. Shapiro. Solving chance-constrained stochastic programs via sampling and integer programming. In *State-of-the-art decision-making tools in the information-intensive age*, pages 261–269. Informs, 2008.
3. R. Atawia, H. Abou-zeid, H. S. Hassanein, and A. Noureldin. Chance-constrained qos satisfaction for predictive video streaming. In *2015 IEEE 40th Conference on Local Computer Networks (LCN)*, pages 253–260. IEEE, 2015.
4. R. Atawia, H. Abou-Zeid, H. S. Hassanein, and A. Noureldin. Joint chance-constrained predictive resource allocation for energy-efficient video streaming. *IEEE Journal on Selected Areas in Communications*, 34(5):1389–1404, 2016.
5. R. Atawia, H. S. Hassanein, H. Abou-Zeid, and A. Noureldin. Robust content delivery and uncertainty tracking in predictive wireless networks. *IEEE Transactions on Wireless Communications*, 16(4):2327–2339, 2017.
6. R. Atawia, H. S. Hassanein, N. A. Ali, and A. Noureldin. Utilization of stochastic modeling for green predictive video delivery under network uncertainties. *IEEE Transactions on Green Communications and Networking*, 2(2):556–569, 2018.
7. J. Blanchet and K. Murthy. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600, 2019.

8. Y. Cai, G. Huang, Z. Yang, Q. Lin, and Q. Tan. Community-scale renewable energy systems planning under uncertainty—an interval chance-constrained programming approach. *Renewable and Sustainable Energy Reviews*, 13(4):721–735, 2009.
9. G. C. Calafiore and M. C. Campi. The scenario approach to robust control design. *IEEE Transactions on automatic control*, 51(5):742–753, 2006.
10. G. C. Calafiore and L. E. Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.
11. Z. Chen, D. Kuhn, and W. Wiesemann. Data-driven chance constrained programs over wasserstein balls. *Operations Research*, 2022.
12. Z. Chen, D. Kuhn, and W. Wiesemann. On Approximations of Data-Driven Chance Constrained Programs over Wasserstein Balls. *arXiv preprint arXiv:2206.00231*, 2022.
13. Z. Chen and W. Xie. Sharing the value-at-risk under distributional ambiguity. *Mathematical Finance*, 31(1):531–559, 2021.
14. J. Cheng, E. Delage, and A. Lisser. Distributionally robust stochastic knapsack problem. *SIAM Journal on Optimization*, 24(3):1485–1506, 2014.
15. R. Gao and A. Kleywegt. Distributionally robust stochastic optimization with wasserstein distance. *Mathematics of Operations Research*, 2022.
16. L. E. Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations research*, 51(4):543–556, 2003.
17. G. Guillén-Gosálbez and I. Grossmann. A global optimization strategy for the environmentally conscious design of chemical supply chains under uncertainty in the damage assessment model. *Computers & Chemical Engineering*, 34(1):42–58, 2010.
18. A. Gupta, C. D. Maranas, and C. M. McDonald. Mid-term supply chain planning under demand uncertainty: customer demand satisfaction and inventory management. *Computers & Chemical Engineering*, 24(12):2613–2621, 2000.
19. A. K. Gupta, T. Varga, and T. Bodnar. *Elliptically contoured models in statistics and portfolio theory*. Springer, 2013.
20. G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. A Distributionally Robust Perspective on Uncertainty Quantification and Chance Constrained Programming. *Mathematical Programming*, 151(1):35–62, 2015.
21. G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, 65(3):751–767, 2017.
22. N. Ho-Nguyen, F. Kılınç-Karzan, S. Küçükyavuz, and D. Lee. Distributionally robust chance-constrained programs with right-hand side uncertainty under wasserstein ambiguity. *Mathematical Programming*, 196(1):641–672, 2022.
23. N. Ho-Nguyen, F. Kilinç-Karzan, S. Küçükyavuz, and D. Lee. Strong formulations for distributionally robust chance-constrained programs with left-hand side uncertainty under wasserstein ambiguity. *INFORMS Journal on Optimization*, 2022.
24. R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.
25. R. Ji and M. A. Lejeune. Data-driven distributionally robust chance-constrained optimization with wasserstein metric. *Journal of Global Optimization*, 79(4):779–811, 2021.
26. N. Jiang and W. Xie. ALSO-X and ALSO-X+: Better Convex Approximations for Chance Constrained Programs. *Operations Research*, 70(6):3581–3600, 2022.
27. N. Jiang and W. Xie. Dfo: A framework for data-driven decision-making with endogenous outliers. 2022.
28. S. Ken-Iti. *Lévy processes and infinitely divisible distributions*. Cambridge university press, 1999.
29. S. Küçükyavuz and R. Jiang. Chance-constrained optimization under limited distributional information: a review of reformulations based on sampling and distributional robustness. *EURO Journal on Computational Optimization*, page 100030, 2022.
30. C. M. Lagoa, X. Li, and M. Sznaier. Probabilistically constrained linear programs and risk-adjusted controller design. *SIAM Journal on Optimization*, 15(3):938–951, 2005.

31. M. A. Lejeune and F. Margot. Solving chance-constrained optimization problems with stochastic quadratic inequalities. *Operations Research*, 64(4):939–957, 2016.
32. B. Li, R. Jiang, and J. L. Mathieu. Ambiguous risk constraints with moment and unimodality information. *Mathematical Programming*, 173:151–192, 2019.
33. S. Li, Y. T. Hou, W. Lou, B. A. Jalaian, and S. Russell. Maximizing energy efficiency with channel uncertainty under mutual interference. *IEEE Transactions on Wireless Communications*, 21(10):8476–8488, 2022.
34. S. Li, Y. Huang, C. Li, B. A. Jalaian, Y. T. Hou, and W. Lou. Coping uncertainty in coexistence via exploitation of interference threshold violation. In *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 71–80, 2019.
35. A. Nemirovski and A. Shapiro. Scenario approximations of chance constraints. *Probabilistic and randomized methods for design under uncertainty*, pages 3–47, 2006.
36. A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2007.
37. B. K. Pagnoncelli, S. Ahmed, and A. Shapiro. Computational study of a chance constrained portfolio selection problem. *Journal of Optimization Theory and Applications*, 142(2):399–416, 2009.
38. B. K. Pagnoncelli, D. Reich, and M. C. Campi. Risk-return trade-off with the scenario approach in practice: a case study in portfolio selection. *Journal of Optimization Theory and Applications*, 155(2):707–722, 2012.
39. A. Prékopa. Programming under probabilistic constraints with a random technology matrix. *Statistics: A Journal of Theoretical and Applied Statistics*, 5(2):109–116, 1974.
40. A. Prékopa. *Stochastic programming*, volume 324. Springer Science & Business Media, 2013.
41. R. T. Rockafellar, S. Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.
42. R. T. Rockafellar and R. J. Wets. On the interchange of subdifferentiation and conditional expectation for convex functionals. *Stochastics: An International Journal of Probability and Stochastic Processes*, 7(3):173–182, 1982.
43. A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.
44. H. Shen and R. Jiang. Convex chance-constrained programs with wasserstein ambiguity. *arXiv preprint arXiv:2111.02486*, 2021.
45. H. Shen and R. Jiang. Chance-constrained set covering with wasserstein ambiguity. *Mathematical Programming*, pages 1–54, 2022.
46. Z. Shi, H. Liang, S. Huang, and V. Dinavahi. Distributionally robust chance-constrained energy management for islanded microgrids. *IEEE Transactions on Smart Grid*, 10(2):2234–2244, 2018.
47. S. Uryasev. Introduction to the theory of probabilistic functions and percentiles (value-at-risk). In *Probabilistic Constrained Optimization*, pages 1–25. Springer, 2000.
48. K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi. Outage constrained robust transmit optimization for multiuser miso downlinks: Tractable approximations by conic optimization. *IEEE Transactions on Signal Processing*, 62(21):5690–5705, 2014.
49. W. Xie. On distributionally robust chance constrained programs with wasserstein distance. *Mathematical Programming*, 186(1):115–155, 2021.
50. W. Xie and S. Ahmed. On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM Journal on Optimization*, 28(2):1151–1182, 2018.
51. W. Xie and S. Ahmed. Bicriteria approximation of chance-constrained covering problems. *Operations Research*, 68(2):516–533, 2020.
52. J. Yeh. *Real analysis: theory of measure and integration second edition*. World Scientific Publishing Company, 2006.
53. M.-C. Yue, D. Kuhn, and W. Wiesemann. On linear optimization over wasserstein balls. *Mathematical Programming*, 195(1):1107–1122, 2022.

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

A Proofs

A.1 Proof of Theorem 5

Theorem 5 *Suppose that in a single DRCCP* (2)*, the deterministic set $\mathcal{X}$ is convex, the reference distribution $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ is continuous and nondegenerate with support $\mathbb{R}^n$, $\|\cdot\|_* = \|\cdot\|_p$ with $p \in (1,\infty)$ and affine mappings $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$ and $b_1(\boldsymbol{x}) = b_1$. Then the lower-level* ALSO-X (3b) *admits a unique optimal solution when $t \geq \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and $v^A(t) > 0$.*

Proof We split the proof into two parts by discussing whether there exists an optimal solution $\boldsymbol{x}^*$ to the lower-level ALSO-X (3b) such that $\mathrm{supp}(\boldsymbol{x}^*) = n$.

(i) Suppose that there exists an optimal solution $\boldsymbol{x}^*$ to the lower-level ALSO-X such that $\mathrm{supp}(\boldsymbol{x}^*) = n$. First, for any $\boldsymbol{x}$ with $\mathrm{supp}(\boldsymbol{x}) = n$, according to the continuity of function $f(\tau) = \max\{\tau, 0\}$ and theorem 1 in [42], we can interchange the subdifferential operator and expectation, the first-order derivative $\widehat{F}'(\boldsymbol{x})$ is

$$\widehat{F}'(\boldsymbol{x}) = \frac{\partial \widehat{F}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \int_{\mu(\boldsymbol{x})} \frac{\partial}{\partial \boldsymbol{x}} \left[\theta \left\|\boldsymbol{x}\right\|_* + \boldsymbol{x}^\top \boldsymbol{\zeta}\right] \mathbb{P}(d\boldsymbol{\zeta}),$$

where $\mu(\boldsymbol{x}) = \{\boldsymbol{\zeta} : \theta\|\boldsymbol{x}\|_* + \boldsymbol{x}^\top\boldsymbol{\zeta} \geq b_1\}$ with its boundary $\partial\mu(\boldsymbol{x}) = \{\boldsymbol{\zeta} : \theta\|\boldsymbol{x}\|_* + \boldsymbol{x}^\top\boldsymbol{\zeta} = b_1\}$. According to equation (4) in [47], the Hessian of $\widehat{F}(\boldsymbol{x})$ is

$$H_{\widehat{F}}(\boldsymbol{x}) = \frac{1}{\|\boldsymbol{x}\|_2} \int_{\partial\mu(\boldsymbol{x})} \left(\frac{\partial\theta\left\|\boldsymbol{x}\right\|_*}{\partial\boldsymbol{x}} + \boldsymbol{\zeta}\right)\left(\frac{\partial\theta\left\|\boldsymbol{x}\right\|_*}{\partial\boldsymbol{x}} + \boldsymbol{\zeta}\right)^\top \mathbb{P}(d\boldsymbol{\zeta}) + \theta \int_{\mu(\boldsymbol{x})} \frac{\partial^2\|\boldsymbol{x}\|_*}{\partial\boldsymbol{x}^2}\mathbb{P}(d\boldsymbol{\zeta}).$$

Recall that $\boldsymbol{x}^*$ is optimal to the lower-level ALSO-X (3b) and $\boldsymbol{x}^* \neq \boldsymbol{0}$. To show that $\boldsymbol{x}^*$ is the unique optimal solution, it suffices to show that its corresponding Hessian $H_{\widehat{F}}(\boldsymbol{x})$ is positive definite (PD). For any $\boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{y} \neq \boldsymbol{0}$, we have

$$\boldsymbol{y}^\top H_{\widehat{F}}(\boldsymbol{x})\boldsymbol{y} = \frac{1}{\|\boldsymbol{x}\|_2} \int_{\partial\mu(\boldsymbol{x})} \left(\boldsymbol{y}^\top \frac{\partial\theta\left\|\boldsymbol{x}\right\|_*}{\partial\boldsymbol{x}} + \boldsymbol{y}^\top\boldsymbol{\zeta}\right)^2 \mathbb{P}(d\boldsymbol{\zeta}) + \theta\int_{\mu(\boldsymbol{x})} \boldsymbol{y}^\top \left(\frac{\partial^2\|\boldsymbol{x}\|_*}{\partial\boldsymbol{x}^2}\right)\boldsymbol{y}\mathbb{P}(d\boldsymbol{\zeta}).$$

For $\|\boldsymbol{x}\|_* = \|\boldsymbol{x}\|_p$ with $p \in (1,\infty)$, for each $i \in [n]$, we have

$$\frac{\partial\|\boldsymbol{x}\|_p}{\partial x_i} = \mathrm{sign}(x_i)|x_i|^{p-1}\left[\sum_{j\in[n]} |x_j|^p\right]^{\frac{1}{p}-1}, \forall i \in [n],$$

and

$$\frac{\partial^2\|\boldsymbol{x}\|_p}{\partial x_i^2} = |x_i|^{p-2}\left[\sum_{j\in[n]\setminus\{i\}} |x_j|^p\right](p-1)\left[\sum_{j\in[n]} |x_j|^p\right]^{\frac{1}{p}-2}, \forall i \in [n],$$

$$\frac{\partial^2\|\boldsymbol{x}\|_p}{\partial x_i \partial x_k} = \mathrm{sign}(x_i)|x_i|^{p-1}\,\mathrm{sign}(x_k)|x_k|^{p-1}(1-p)\left[\sum_{j\in[n]} |x_j|^p\right]^{\frac{1}{p}-2}, \forall i \in [n], k \in [n]\setminus\{i\}.$$

That is,

$$\frac{\partial^2\|\boldsymbol{x}\|_*}{\partial\boldsymbol{x}^2} = \frac{\partial^2\|\boldsymbol{x}\|_p}{\partial\boldsymbol{x}^2} = (p-1)\left(\sum_{j\in[n]}|x_j|^p\right)^{\frac{1}{p}-2}\left[\left(\sum_{j\in[n]}|x_j|^p\right)\mathrm{Diag}\begin{pmatrix}|x_1|^{p-2}\\ \vdots \\ |x_n|^{p-2}\end{pmatrix}\right.$$

$$
- \begin{pmatrix} \text{sign}(x_1)|x_1|^{p-1} \\ \vdots \\ \text{sign}(x_n)|x_n|^{p-1} \end{pmatrix} \begin{pmatrix} \text{sign}(x_1)|x_1|^{p-1} \\ \vdots \\ \text{sign}(x_n)|x_n|^{p-1} \end{pmatrix}^{\top} \Bigg] .
$$

Here, $\partial^2\|\boldsymbol{x}\|_p/\partial\boldsymbol{x}^2$ is a positive semidefinite (PSD) matrix of rank $n-1$. This implies that (i) the value $\boldsymbol{y}^\top(\partial^2\|\boldsymbol{x}\|_*/\partial\boldsymbol{x}^2)\boldsymbol{y} = 0$ if and only if $\boldsymbol{y} = \tau\boldsymbol{x}$ with $\tau \neq 0$ (recall that we have $\boldsymbol{y} \neq \mathbf{0}$); and (ii) if $\boldsymbol{z}$ is an alternative optimal solution, then we must have $\boldsymbol{z} = \ell\boldsymbol{x}^*$. Hence, if $\boldsymbol{y} \not\propto \boldsymbol{x}^*$, then we must have $\boldsymbol{y}^\top(\partial^2\|\boldsymbol{x}^*\|_*/\partial\boldsymbol{x}^2)\boldsymbol{y} > 0$, which implies that the Hessian $H_{\widehat{F}}(\boldsymbol{x}^*)$ is PD, which confirms the uniqueness of the optimal solution $\boldsymbol{x}^*$. It remains to show that if the solution $\boldsymbol{x}^*$ is not unique, all optimal solutions should be proportional to $\boldsymbol{x}^*$. We split the proof into two steps by the sign of b_1.

Step I. When $b_1 \neq 0$, we show that the Hessian $H_{\widehat{F}}(\boldsymbol{x}^*)$ is PD. That is, it suffices to show that if $\boldsymbol{y} = \ell\boldsymbol{x}^*$ with $\ell \neq 0$, we must have

$$
\int_{\partial\mu(\boldsymbol{x}^*)} \left(\boldsymbol{y}^\top \frac{\partial\theta\,\|\boldsymbol{x}^*\|_*}{\partial\boldsymbol{x}} + \boldsymbol{y}^\top\boldsymbol{\zeta} \right)^2 \mathbb{P}(d\boldsymbol{\zeta}) > 0.
$$

Indeed, we notice that

$$
\boldsymbol{y}^\top \frac{\partial\theta\,\|\boldsymbol{x}^*\|_*}{\partial\boldsymbol{x}} + \boldsymbol{y}^\top\boldsymbol{\zeta} = (\ell\boldsymbol{x}^*)^\top \frac{\partial\theta\,\|\boldsymbol{x}^*\|_*}{\partial\boldsymbol{x}} + (\ell\boldsymbol{x}^*)^\top\boldsymbol{\zeta} = \ell\left[\theta\,\|\boldsymbol{x}^*\|_* + (\boldsymbol{x}^*)^\top\boldsymbol{\zeta}\right] = \ell b_1 \neq 0.
$$

Hence,

$$
\int_{\partial\mu(\boldsymbol{x}^*)} \left(\boldsymbol{y}^\top \frac{\partial\theta\,\|\boldsymbol{x}^*\|_*}{\partial\boldsymbol{x}} + \boldsymbol{y}^\top\boldsymbol{\zeta} \right)^2 \mathbb{P}(d\boldsymbol{\zeta}) = \int_{\partial\mu(\boldsymbol{x}^*)} \ell^2 b_1^2 \mathbb{P}(d\boldsymbol{\zeta}) > 0,
$$

which implies that $\boldsymbol{y}^\top H_{\widehat{F}}(\boldsymbol{x}^*)\boldsymbol{y} > 0$.

Step II. When $b_1 = 0$, we show that $\boldsymbol{x}^*$ must be the unique solution. Suppose that there exists another optimal solution $\boldsymbol{z}$. Then we must have $\boldsymbol{z} = \ell\boldsymbol{x}^*$ such that $\ell \neq 1$ according to the statement at the beginning of the proof. Three subcases remain to be discussed:

Subcase (i) When $0 < \ell < 1$, according to the homogeneity of the objective function in the lower-level ALSO-X (3b), i.e.,

$$
\mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}}\left[\theta\,\|\boldsymbol{z}\|_* + \boldsymbol{z}^\top\tilde{\boldsymbol{\zeta}}\right]_+ = \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}}\left[\theta\,\|\ell\boldsymbol{x}^*\|_* + (\ell\boldsymbol{x}^*)^\top\tilde{\boldsymbol{\zeta}}\right]_+ = \ell v^A(t).
$$

Hence, $\boldsymbol{z} = \ell\boldsymbol{x}^*$ yields a strict better objective value than that of $\boldsymbol{x}^*$ as $v^A(t) > 0$, a contradiction;

Subcase (ii) Similarly, when $\ell > 1$, the current optimal solution $\boldsymbol{x}^*$ is strict better than $\boldsymbol{z}$, a contradiction;

Subcase (iii) When $\ell < 0$, since the objective function in the lower-level ALSO-X (3b) and the feasible region is convex, any convex combination of $\boldsymbol{x}^*$ and $\boldsymbol{z} = \ell\boldsymbol{x}^*$ is also an optimal solution. That is, if we choose

$$
\frac{-\ell}{|\ell|+1}\boldsymbol{x}^* + \frac{1}{|\ell|+1}(\ell\boldsymbol{x}^*) = \mathbf{0},
$$

then $\mathbf{0}$ is one optimal solution with objective value 0, which is strictly less than that of $\boldsymbol{x}^*$, a contradiction. Thus, we have that $\boldsymbol{x}^*$ with $|\operatorname{supp}(\boldsymbol{x}^*)| = n$ is the unique optimal solution since we have $v^A(t) > 0$.

Hence, the objective function of the lower-level ALSO-X(3b) admits a unique solution.

(ii) Suppose that there does not exist an optimal solution $\boldsymbol{x}^*$ to the lower-level ALSO-X (3b) such that $|\operatorname{supp}(\boldsymbol{x}^*)| = n$. Let $\boldsymbol{x}^*$ be an optimal solution that has the largest support. If $\boldsymbol{z}$ is another optimal solution to the lower-level ALSO-X (3b) and $\operatorname{supp}(\boldsymbol{x}^*) = \operatorname{supp}(\boldsymbol{z})$, then following the proof in Part (i), we must have $\boldsymbol{x}^* = \boldsymbol{z}$. Thus, $\operatorname{supp}(\boldsymbol{x}^*) \neq \operatorname{supp}(\boldsymbol{z})$. Now let us define

$$\gamma = \frac{1}{2}\frac{\min\{|x_i^*| : i \in \operatorname{supp}(\boldsymbol{x}^*)\}}{\min\{|x_i^*| : i \in \operatorname{supp}(\boldsymbol{x}^*)\} + \max\{|z_i| : i \in \operatorname{supp}(\boldsymbol{z})\}}.$$

Then $\bar{\boldsymbol{x}} := (1-\gamma)\boldsymbol{x}^* + \gamma \boldsymbol{z}$ is another optimal solution, according to the convexity of the feasible set of the lower-level ALSO-X. However, $|\operatorname{supp}(\bar{\boldsymbol{x}})| \geq |\operatorname{supp}(\boldsymbol{x}^*)| + 1$, a contradiction that $\boldsymbol{x}^*$ has the largest support. This completes the proof. □

A.2 Proof of Theorem 5 with $q \in \{1, \infty\}$

Lemma 1 *Suppose that in a single DRCCP* (2), *the deterministic set* $\mathcal{X}$ *is convex, the reference distribution* $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ *is continuous with support* $\mathbb{R}^n$, $\|\cdot\|_* = \|\cdot\|_p$ *with* $p \in \{1, \infty\}$ *and affine mappings* $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$ *and* $b_1(\boldsymbol{x}) = b_1$. *Then the lower-level* ALSO-X (3b) *admits a unique optimal solution when* $t \geq \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ *and* $v^A(t) > 0$.

Proof We split the proof into two cases based on the value of p.
Case I. $p = 1$**.** We split the proof into two parts by discussing whether there exists an optimal solution $\boldsymbol{x}^*$ to the lower-level ALSO-X (3b) such that $\operatorname{supp}(\boldsymbol{x}^*) = n$.

(i) Suppose that there exists an optimal solution $\boldsymbol{x}^*$ to the lower-level ALSO-X such that $\operatorname{supp}(\boldsymbol{x}^*) = n$. Recall that $\widehat{F}(\boldsymbol{x})$ denotes the objective function in the lower-level ALSO-X (3b) with the Hessian

$$H_{\widehat{F}}(\boldsymbol{x}) = \frac{1}{\|\boldsymbol{x}\|_2}\int_{\partial\mu(\boldsymbol{x})}\left(\frac{\partial\theta\,\|\boldsymbol{x}\|_1}{\partial\boldsymbol{x}} + \boldsymbol{\zeta}\right)\left(\frac{\partial\theta\,\|\boldsymbol{x}\|_1}{\partial\boldsymbol{x}} + \boldsymbol{\zeta}\right)^\top \mathbb{P}(d\boldsymbol{\zeta}) + \theta\int_{\mu(\boldsymbol{x})}\frac{\partial^2\|\boldsymbol{x}\|_1}{\partial\boldsymbol{x}^2}\mathbb{P}(d\boldsymbol{\zeta}).$$

For function $\|\boldsymbol{x}\|_1$ with $\operatorname{supp}(\boldsymbol{x}) = n$, we have

$$\frac{\partial\|\boldsymbol{x}\|_1}{\partial x_i} = \operatorname{sign}(x_i), \forall i \in [n]; \frac{\partial^2\|\boldsymbol{x}\|_1}{\partial x_i \partial x_j} = 0, \forall i \in [n], j \in [n].$$

Here, with the assumption that $\operatorname{supp}(\boldsymbol{x}^*) = n$, we have $\partial^2\|\boldsymbol{x}^*\|_1/\partial\boldsymbol{x}^2 = 0$, which implies that $\boldsymbol{y}^\top(\partial^2\|\boldsymbol{x}\|_1/\partial\boldsymbol{x}^2)\boldsymbol{y} = 0$. Following the similar proof of Theorem 5, we have

$$\int_{\partial\mu(\boldsymbol{x}^*)}\left(\boldsymbol{y}^\top\frac{\partial\theta\,\|\boldsymbol{x}^*\|_1}{\partial\boldsymbol{x}} + \boldsymbol{y}^\top\boldsymbol{\zeta}\right)^2 \mathbb{P}(d\boldsymbol{\zeta}) > 0.$$

(ii) The case when $|\operatorname{supp}(\boldsymbol{x}^*)| < n$ is identical to part (ii) in the proof of Theorem 5 and is thus omitted.

Case II. $p = \infty$**.** We split the proof into two parts by discussing whether there exists an optimal solution $\boldsymbol{x}^*$ to the lower-level ALSO-X (3b) such that $\operatorname{supp}(\boldsymbol{x}^*) = n$.

(i) Suppose that there exists an optimal solution $\boldsymbol{x}^*$ to the lower-level ALSO-X such that $\operatorname{supp}(\boldsymbol{x}^*) = n$. We split the following proof into two steps.
Step I. We denote $\mathcal{S}$ to be an index set corresponding to the largest absolute-value component, that is, for any subset $\mathcal{S} \subseteq [n]$, we assume $|x_i^*| = \|\boldsymbol{x}^*\|_\infty$ for each $i \in \mathcal{S}$. For the fixed subset $\mathcal{S}$ and function $\|\boldsymbol{x}\|_\infty$ with $\operatorname{supp}(\boldsymbol{x}) = n$, we have

$$\frac{\partial\|\boldsymbol{x}\|_\infty}{\partial x_i} = \begin{cases}\operatorname{sign}(x_i), & i \in \mathcal{S}, \\ 0, & i \notin \mathcal{S}\end{cases}, \forall i \in [n],$$

and

$$\frac{\partial^2 \|\boldsymbol{x}\|_\infty}{\partial x_i \partial x_j} = 0, \forall i \in [n], k \in [n].$$

Here, with the presumptions, we have $\partial^2 \|\boldsymbol{x}^*\|_\infty / \partial \boldsymbol{x}^2 = 0$, which implies that $\boldsymbol{y}^\top (\partial^2 \|\boldsymbol{x}\|_\infty / \partial \boldsymbol{x}^2) \boldsymbol{y} = 0$. Following the similar proof of Theorem 5, we have

$$\int_{\partial \mu(\boldsymbol{x}^*)} \left(\boldsymbol{y}^\top \frac{\partial \theta \, \|\boldsymbol{x}^*\|_\infty}{\partial \boldsymbol{x}} + \boldsymbol{y}^\top \boldsymbol{\zeta} \right)^2 \mathbb{P}(d\boldsymbol{\zeta}) > 0.$$

Thus, given a subset $\mathcal{S}$, the optimal solution is unique.

Step II. It remains to prove that for any subset $\mathcal{S}$, the optimal solution is unique, where $\mathcal{S}$ is an index set corresponding to the largest absolute-value components of $\boldsymbol{x}^*$. Suppose that $\boldsymbol{x}^*$ has the smallest size of the largest absolute-value components. Assume $\boldsymbol{z} \neq \boldsymbol{x}^*$ is another optimal solution to the lower-level ALSO-X (3b) with $\mathcal{S}_1$ being its corresponding index set with the largest absolute-value components. According to our assumption, we must have $|\mathcal{S}_1| \geq |\mathcal{S}|$. Now let us define

$$\gamma = \frac{1}{2} \frac{\min\{|x_i^*| : i \in \operatorname{supp}(\boldsymbol{x}^*)\}}{\min\{|x_i^*| : i \in \operatorname{supp}(\boldsymbol{x}^*)\} + \max\{|z_i^*| : i \in \operatorname{supp}(\boldsymbol{z}^*)\}}.$$

Then the convexity of the feasible set of the lower-level ALSO-X implies that $\bar{\boldsymbol{x}} := (1-\gamma)\boldsymbol{x}^* + \gamma \boldsymbol{z}^*$ is another optimal solution. However, the new solution $\bar{\boldsymbol{x}}$ either $\boldsymbol{x}^*$ is the unique solution of the lower-level ALSO-X or contradicts that $\boldsymbol{x}^*$ has the smallest size of the largest absolute-value components.

(ii) The case when $|\operatorname{supp}(\boldsymbol{x}^*)| < n$ is identical to part (ii) in the proof of Theorem 5 and is thus omitted. □

B Equivalent Reformulations for DRCCPs under Type $q-$Wasserstein Ambiguity Set

Under type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$, the DRCCP (1) can be written as

$$v_q^* = \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \inf_{\mathbb{P} \in \mathcal{P}_q} \mathbb{P} \left\{ \tilde{\boldsymbol{\xi}} \colon \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} \leq b_i(\boldsymbol{x}), \forall i \in [I] \right\} \geq 1 - \varepsilon \right\}. \tag{22}$$

For the notational convenience, we denote the decision space induced by the worst-case chance constraint in DRCCP (22) as the following Distributionally Robust Chance Constrained (DRCC) set

$$Z_q \colon = \left\{ \boldsymbol{x} \in \mathbb{R}^n \colon \inf_{\mathbb{P} \in \mathcal{P}_q} \mathbb{P} \left\{ \tilde{\boldsymbol{\xi}} \colon \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} \leq b_i(\boldsymbol{x}), \forall i \in [I] \right\} \geq 1 - \varepsilon \right\}. \tag{23}$$

B.1 Equivalent Reformulations of DRCCPs

We can generalize the existing work in [49] on the equivalent formulation of DRCC set Z_q (23) for any $q \in [1, \infty)$ and any reference distribution $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$.

Proposition 3 *(A generalization of corollary 1 in [49]) Under type $q-$Wasserstein ambiguity set, DRCC set Z_q* (23) *is equivalent to*

$$Z_q = \left\{ \boldsymbol{x} \in \mathbb{R}^n \colon \begin{array}{l} \theta^q \varepsilon^{-1} + \operatorname{CVaR}_{1-\varepsilon}\left[-f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}})^q \right] \leq 0, \\ \theta^q \varepsilon^{-1} + \operatorname{VaR}_{1-\varepsilon}\left[-f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}})^q \right] \leq 0 \end{array} \right\}. \tag{24}$$

where

$$f(\boldsymbol{x}, \boldsymbol{\zeta}) = \min \left\{ \min_{i \in [I] \setminus \mathcal{I}(\boldsymbol{x})} \frac{(b_i(\boldsymbol{x}) - \boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\zeta})_+}{\|\boldsymbol{a}_i(\boldsymbol{x})\|_*}, \min_{i \in \mathcal{I}(\boldsymbol{x})} \chi_{\{\boldsymbol{x} \colon b_i(\boldsymbol{x}) < 0\}}(\boldsymbol{x}) \right\},$$

and $\mathcal{I}(\boldsymbol{x}) = \emptyset$ if $\boldsymbol{a}_i(\boldsymbol{x}) \neq \boldsymbol{0}$ and $\mathcal{I}(\boldsymbol{x}) = [I]$.

Proof We split our proof into two cases by discussing whether $\theta = 0$ or not.

Case 1 When $\theta = 0$, DRCC set Z_q reduces to the regular CCP under reference distribution $\mathbb{P}_{\tilde{\zeta}}$, i.e.,

$$Z_q = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \mathbb{P}\left\{ \tilde{\boldsymbol{\zeta}} \colon \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) \leq 0, \forall i \in [I] \right\} \geq 1 - \varepsilon \right\}.$$

On the other hand, in (24), the first constraint is redundant when $\theta = 0$. Hence, the statement follows.

Case 2 When $\theta > 0$, the fact that the decision space induced by the first constraint in (24) is equivalent to DRCC set Z_q follows directly from the proof of corollary 1 in [49]. Thus, in this case, the second constraint in (24) is redundant since for any random variable $\tilde{\boldsymbol{X}}$, we have $\mathrm{CVaR}_{1-\varepsilon}[\tilde{\boldsymbol{X}}] \geq \mathrm{VaR}_{1-\varepsilon}[\tilde{\boldsymbol{X}}]$.

□

The reformulation in Proposition 3 can be simplified if $\|\boldsymbol{a}_i(\boldsymbol{x})\|_* = \|\boldsymbol{a}_1(\boldsymbol{x})\|_*$ for all $i \in [I]$, in which the condition can be viewed as a generalization of $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{a}_i(\boldsymbol{x})$ for all $i \in [I]$, which has been discussed in the DRCCP literature (see, e.g., [11, 49]). Notice that this condition always holds for a single DRCCP, where $I = 1$.

Proposition 4 *Suppose that $\|\boldsymbol{a}_i(\boldsymbol{x})\|_* = \|\boldsymbol{a}_1(\boldsymbol{x})\|_*$ for all $i \in [I]$. Then DRCC set Z_q* (23) *can be simplified to*

$$Z_q = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{c} \theta^q \varepsilon^{-1} \|\boldsymbol{a}_1(\boldsymbol{x})\|_*^q + \mathrm{CVaR}_{1-\varepsilon}\left[-\min_{i \in [I]} \left(b_i(\boldsymbol{x}) - \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} \right)_+^q \right] \leq 0, \\ \mathbb{P}\left\{ \tilde{\boldsymbol{\zeta}} \colon \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) \leq 0, \forall i \in [I] \right\} \geq 1 - \varepsilon \end{array} \right\}. \tag{25}$$

Proof We split our proof into two cases by discussing whether $\|\boldsymbol{a}_1(\boldsymbol{x})\|_* = 0$ or not.

Case 1 When $\|\boldsymbol{a}_1(\boldsymbol{x})\|_* = 0$ (i.e., $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{0}$), according to (24), set $Z_q \cap \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{a}_1(\boldsymbol{x})\|_* = 0\}$ is equivalent to the set

$$\{\boldsymbol{x} \in \mathbb{R}^n \colon \|\boldsymbol{a}_1(\boldsymbol{x})\|_* = 0, b_i(\boldsymbol{x}) \geq 0, \forall i \in [I]\},$$

which is equivalent to the right-hand side of (25) intersecting with set $\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{a}_1(\boldsymbol{x})\|_* = 0\}$ since its first constraint is redundant.

Case 2 When $\|\boldsymbol{a}_1(\boldsymbol{x})\|_* > 0$, according to Proposition 3, the function $f(\boldsymbol{x}, \boldsymbol{\zeta})$ becomes

$$f(\boldsymbol{x}, \boldsymbol{\zeta}) = \min_{i \in [I]} \frac{\left(b_i(\boldsymbol{x}) - \boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\zeta} \right)_+}{\|\boldsymbol{a}_1(\boldsymbol{x})\|_*}.$$

According to the positive homogeneity of the coherent risk measure CVaR, set $Z_q \cap \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{a}_1(\boldsymbol{x})\|_* > 0\}$ is equivalent to the set

$$\left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{c} \|\boldsymbol{a}_1(\boldsymbol{x})\|_* > 0, \\ \theta^q \varepsilon^{-1} \|\boldsymbol{a}_1(\boldsymbol{x})\|_*^q + \mathrm{CVaR}_{1-\varepsilon}\left[-\min_{i \in [I]} \left(b_i(\boldsymbol{x}) - \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} \right)_+^q \right] \leq 0, \\ \mathbb{P}\left\{ \tilde{\boldsymbol{\zeta}} \colon \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) \leq 0, \forall i \in [I] \right\} \geq 1 - \varepsilon \end{array} \right\}.$$

This completes the proof. □

As a direct corollary of Proposition 4, we remark that for the single DRCCP (i.e., $I = 1$) with the elliptical reference distribution $\mathbb{P}_{\tilde{\zeta}}$ (see the discussions in Section 5.1), DRCC set Z_q admits a simple representation.

Corollary 11 *For the single DRCCP* (22)*, when the affine mappings are $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$, $b_1(\boldsymbol{x}) = b_1$, the random parameters $\tilde{\boldsymbol{\zeta}}$ follow a joint elliptical distribution with $\tilde{\boldsymbol{\zeta}} \sim \mathbb{P}_{\mathrm{E}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \widehat{g})$, and the norm defining the Wasserstein distance is the generalized Mahalanobis norm associated with the matrix $\boldsymbol{\Sigma}$, DRCC set Z_q* (23) *becomes*

$$Z_q = \left\{ \boldsymbol{x} \in \mathbb{R}^n \colon b_1(\boldsymbol{x}) - \boldsymbol{\mu}^\top \boldsymbol{a}_1(\boldsymbol{x}) \geq \eta_q^* \sqrt{\boldsymbol{a}_1(\boldsymbol{x})^\top \boldsymbol{\Sigma} \boldsymbol{a}_1(\boldsymbol{x})} \right\}, \tag{26a}$$

and η_q^* is the unique minimizer of

$$\eta_q^* = \min_{\eta} \left\{ \eta \colon \int_{\Phi^{-1}(1-\varepsilon)}^{\eta} (\eta - t)^q \bar{k} \widehat{g}(t^2/2) dt \geq \theta^q, \eta \geq \Phi^{-1}(1-\varepsilon) \right\}. \tag{26b}$$

Proof We note that if $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{0}$, according to Proposition 4, set $Z_q \cap \{\boldsymbol{x} \in \mathbb{R}^n \colon \boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{0}\}$ becomes

$$\{\boldsymbol{x} \in \mathbb{R}^n \colon \boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{0}, b_1(\boldsymbol{x}) \geq 0\},$$

which is equivalent to the right-hand side of (26a) intersecting with set $\{\boldsymbol{x} \in \mathbb{R}^n \colon \boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{0}\}$. Thus, without loss of generality, we assume that $\boldsymbol{a}_1(\boldsymbol{x}) \neq \boldsymbol{0}$.

Note that the linear function $\boldsymbol{a}_1(\boldsymbol{x})^\top \boldsymbol{\zeta}$ is still elliptically distributed (see, e.g., [19]). For ease of notation, we denote the distribution of the linear function $\boldsymbol{a}_1(\boldsymbol{x})^\top \boldsymbol{\zeta}$ as $\mathbb{P}_{\mathrm{E}}(\mu_{\boldsymbol{x}}, \sigma_{\boldsymbol{x}}, \widehat{g})$ and denote its probability density function as

$$h(y) = \frac{\bar{k}}{\sigma_{\boldsymbol{x}}} \widehat{g} \left(\frac{(y - \mu_{\boldsymbol{x}})^2}{2\sigma_{\boldsymbol{x}}^2} \right),$$

where $\mu_{\boldsymbol{x}} = \boldsymbol{\mu}^\top \boldsymbol{a}_1(\boldsymbol{x})$ and $\sigma_{\boldsymbol{x}} = \sqrt{\boldsymbol{a}_1(\boldsymbol{x})^\top \boldsymbol{\Sigma} \boldsymbol{a}_1(\boldsymbol{x})}$.

According to Proposition 4, DRCC set Z_q is

$$Z_q = \left\{ \boldsymbol{x} \in \mathbb{R}^n \colon \begin{array}{l} \theta^q \varepsilon^{-1} \|\boldsymbol{a}_1(\boldsymbol{x})\|_*^q + \mathrm{CVaR}_{1-\varepsilon} \left[-\left(b_1(\boldsymbol{x}) - \boldsymbol{a}_1(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} \right)_+^q \right] \leq 0, \\ \mathbb{P}\left\{ \tilde{\boldsymbol{\zeta}} \colon \boldsymbol{a}_1(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_1(\boldsymbol{x}) \leq 0 \right\} \geq 1 - \varepsilon \end{array} \right\}.$$

Following the similar derivation in theorem 7 [13] and according to the definition of CVaR (see, e.g., [41]), set Z_q is equivalent to

$$Z_q = \left\{ \boldsymbol{x} \in \mathbb{R}^n \colon \frac{1}{\varepsilon} \int_{\mathrm{VaR}_{1-\varepsilon}[\boldsymbol{a}_1(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}}]}^{b_1(\boldsymbol{x})} (b_1(\boldsymbol{x}) - y)^q h(y) dy \geq \theta^q \varepsilon^{-1} \|\boldsymbol{a}_1(\boldsymbol{x})\|_*^q, b_1(\boldsymbol{x}) \geq \mathrm{VaR}_{1-\varepsilon} \left[\boldsymbol{a}_1(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} \right] \right\}.$$

According to theorem 3 in [39], we have

$$\mathrm{VaR}_{1-\varepsilon} \left[\boldsymbol{a}_1(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} \right] = \mu_{\boldsymbol{x}} + \Phi^{-1}(1-\varepsilon) \sigma_{\boldsymbol{x}}.$$

Now let $t = (y - \mu_{\boldsymbol{x}})/\sigma_{\boldsymbol{x}}, y = t\sigma_{\boldsymbol{x}} + \mu_{\boldsymbol{x}}$ and $\eta = (b_1(\boldsymbol{x}) - \mu_{\boldsymbol{x}})/\sigma_{\boldsymbol{x}}$, and then DRCC set Z_q is further equal to

$$Z_q = \left\{ \boldsymbol{x} \in \mathbb{R}^n \colon b_1(\boldsymbol{x}) - \mu_{\boldsymbol{x}} \geq \eta \sigma_{\boldsymbol{x}}, \int_{\Phi^{-1}(1-\varepsilon)}^{\eta} (\eta \sigma_{\boldsymbol{x}} - t\sigma_{\boldsymbol{x}})^q h(t\sigma_{\boldsymbol{x}} + \mu_{\boldsymbol{x}}) dt \geq \theta^q \sigma_{\boldsymbol{x}}^{q-1}, \eta \geq \Phi^{-1}(1-\varepsilon) \right\}.$$

where we use the fact that $\|\boldsymbol{a}_1(\boldsymbol{x})\|_* = \sigma_{\boldsymbol{x}}$. We see that set Z_q expands as η decreases, and thus, we can replace it by the minimal η_q^* defined (26b). Substituting the generating function $\widehat{g}(\cdot)$, we arrive at (26a). Finally, we note that $\int_{\Phi^{-1}(1-\varepsilon)}^{\eta} (\eta - t)^q \bar{k} \widehat{g}(t^2/2) dt$ is monotone increasing in η, which demonstrates the uniqueness of η_q^*. □

B.2 Equivalent Reformulations of ALSO-X

Similar to ALSO-X (3), we extend the ALSO-X under type $q-$Wasserstein ambiguity set. For any $q \in [1, \infty)$, ALSO-X is formally defined as

$$v_q^A = \min_t \quad t, \tag{27a}$$

$$\text{s.t.} \quad \boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{P}_q} \left\{ \mathbb{E}_{\mathbb{P}} \left[\max_{i \in [I]} \left(\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} - b_i(\boldsymbol{x}) \right)_+ \right] : \boldsymbol{c}^\top \boldsymbol{x} \le t \right\}, \tag{27b}$$

$$\boldsymbol{x}^* \in Z_q. \tag{27c}$$

We then derive an equivalent reformulation of the hinge-loss approximation (27b) under type $q-$Wasserstein ambiguity set.

Proposition 5 *Under type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$, hinge-loss approximation* (27b) *is equivalent to*

$$v_q^A(t) = \min_{\substack{\boldsymbol{x} \in \mathcal{X}, \\ \lambda \ge 0}} \left\{ \lambda \theta^q + \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left[\max_{i \in [I]} \left(\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) + P_{q,i}(\boldsymbol{x}, \lambda) \right)_+ \right] : \boldsymbol{c}^\top \boldsymbol{x} \le t \right\}, \tag{28}$$

where for each $i \in [I]$, $P_{q,i}(\boldsymbol{x}, \lambda) = (\|\boldsymbol{a}_i(\boldsymbol{x})\|_*)^{\frac{q}{q-1}} \lambda^{-\frac{1}{q-1}} q^{-\frac{q}{q-1}} (q-1)$ with its limit being

$$\lim_{q \to 1_+} P_{q,i}(\boldsymbol{x}, \lambda) = \lim_{q \to 1_+} (\|\boldsymbol{a}_i(\boldsymbol{x})\|_*)^{\frac{q}{q-1}} \lambda^{-\frac{1}{q-1}} q^{-\frac{q}{q-1}} (q-1) = \chi_{\{\mathbb{R}^n : \|\boldsymbol{a}_i(\boldsymbol{x})\|_* \le \lambda\}}(\boldsymbol{x}).$$

Proof According to theorem 1 in [15] or theorem 1 in [7], the inner supremum $\sup_{\mathbb{P} \in \mathcal{P}_q} \mathbb{E}_{\mathbb{P}}[\max_{i \in [I]} (\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} - b_i(\boldsymbol{x}))_+]$ in (27b) can be reformulated as

$$\min_{\boldsymbol{x} \in \mathcal{X}, \lambda \ge 0} \lambda \theta^q - \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left[\inf_{\boldsymbol{\xi}} \left\{ \lambda \left\| \boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}} \right\|^q - \max_{i \in [I]} (\boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\xi} - b_i(\boldsymbol{x}))_+ \right\} \right].$$

Next, we split the proof into two steps.
Step 1. We first reformulate the term $\inf_{\boldsymbol{\xi}} \{\lambda \|\boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}}\|^q - \max_{i \in [I]} (\boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\xi} - b_i(\boldsymbol{x}))_+\}$. Moving the minus sign into the maximum operators, we have

$$\inf_{\boldsymbol{\xi}} \left\{ \min_{i \in [I]} \min \left\{ \lambda \left\| \boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}} \right\|^q - \boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + b_i(\boldsymbol{x}), \lambda \left\| \boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}} \right\|^q \right\} \right\}.$$

Then interchange the minimum and infimum, we obtain

$$\min_{i \in [I]} \min \left\{ \inf_{\boldsymbol{\xi}} \left\{ \lambda \left\| \boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}} \right\|^q - \boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + b_i(\boldsymbol{x}) \right\}, \inf_{\boldsymbol{\xi}} \lambda \left\| \boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}} \right\|^q \right\}.$$

Note that $\inf_{\boldsymbol{\xi}} \lambda \|\boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}}\|^q = 0$ and it remains to simplify $\inf_{\boldsymbol{\xi}} \{\lambda \|\boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}}\|^q - \boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + b_i(\boldsymbol{x})\}$ for each $i \in [I]$. Letting $\widehat{\boldsymbol{\zeta}} = \boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}}$, we have

$$\inf_{\boldsymbol{\xi}} \left\{ \lambda \left\| \boldsymbol{\xi} - \tilde{\boldsymbol{\zeta}} \right\|^q - \boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + b_i(\boldsymbol{x}) \right\} = \inf_{\widehat{\boldsymbol{\zeta}}} \left\{ \lambda \left\| \widehat{\boldsymbol{\zeta}} \right\|^q - \boldsymbol{a}_i(\boldsymbol{x})^\top \widehat{\boldsymbol{\zeta}} \right\} - \boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} + b_i(\boldsymbol{x}).$$

According to Hölder's inequality and the fact that infimum is attainable, we have

$$\begin{aligned} \inf_{\widehat{\boldsymbol{\zeta}}} \left\{ \lambda \left\| \widehat{\boldsymbol{\zeta}} \right\|^q - \boldsymbol{a}_i(\boldsymbol{x})^\top \widehat{\boldsymbol{\zeta}} \right\} &= \inf_{\widehat{\boldsymbol{\zeta}}} \left\{ \lambda \left\| \widehat{\boldsymbol{\zeta}} \right\|^q - \|\boldsymbol{a}_i(\boldsymbol{x})\|_* \left\| \widehat{\boldsymbol{\zeta}} \right\| \right\} \\ &= (\|\boldsymbol{a}_i(\boldsymbol{x})\|_*)^{\frac{q}{q-1}} \lambda^{-\frac{1}{q-1}} q^{-\frac{q}{q-1}} (1-q). \end{aligned}$$

Note that when $q \to 1_+$, the hinge-loss approximation reduces to

$$\lim_{q \to 1_+} (\|\boldsymbol{a}_i(\boldsymbol{x})\|_*)^{\frac{q}{q-1}} \lambda^{-\frac{1}{q-1}} q^{-\frac{q}{q-1}} (q-1) = \chi_{\{\mathbb{R}^n : \|\boldsymbol{a}_i(\boldsymbol{x})\|_* \leq \lambda\}}(\boldsymbol{x}).$$

Step 2. According to Step 1, the hinge-loss approximation (27b) becomes

$$v_q^A(t) = \min_{\boldsymbol{x} \in \mathcal{X}, \lambda \geq 0} \left\{ \lambda \theta^q - \mathbb{E}_{\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}} \left[\min_{i \in [I]} \min \left(-\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} + b_i(\boldsymbol{x}) - (\|\boldsymbol{a}_i(\boldsymbol{x})\|_*)^{\frac{q}{q-1}} \lambda^{-\frac{1}{q-1}} q^{-\frac{q}{q-1}} (q-1), 0 \right) \right] : \boldsymbol{c}^\top \boldsymbol{x} \leq t \right\}.$$

Moving the minus sign inside the expectation, we arrive at the conclusion. □

B.3 Equivalent Reformulations of CVaR Approximation

For DRCCP (22), its CVaR approximation is defined as

$$v_q^{\mathrm{CVaR}} = \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \sup_{\mathbb{P} \in \mathcal{P}_q} \inf_{\beta \leq 0} \left[\beta + \frac{1}{\varepsilon} \mathbb{E}_{\mathbb{P}} \left[\left(\max_{i \in [I]} \left(\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} - b_i(\boldsymbol{x}) \right) - \beta \right)_+ \right] \right] \leq 0 \right\}.$$

Since the ambiguity set P_q is weakly compact according to Assumption A1 and theorem 1 in [53], we can interchange the infimum with the supremum and multiply both sides by ε. Then for any $q \in [1, \infty)$, CVaR approximation can be formulated as

$$v_q^{\mathrm{CVaR}} = \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \inf_{\beta \leq 0} \sup_{\mathbb{P} \in \mathcal{P}_q} \left[\varepsilon\beta + \mathbb{E}_{\mathbb{P}} \left[\left(\max_{i \in [I]} \left(\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}} - b_i(\boldsymbol{x}) \right) - \beta \right)_+ \right] \right] \leq 0 \right\}.$$

Following similar derivations as those of Proposition 3 and Proposition 5, we obtain the equivalent reformulation of CVaR approximation for DRCCP (22).

Proposition 6 *Under type $q-$Wasserstein ambiguity set,* CVaR *approximation of DRCCP* (22) *is equivalent to*

$$v_q^{\mathrm{CVaR}} = \min_{\boldsymbol{x} \in \mathcal{X}, \lambda \geq 0, \beta \leq 0} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \varepsilon\beta + \lambda\theta^q + \mathbb{E}_{\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}} \left[\max_{i \in [I]} \left(\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) + P_{q,i}(\boldsymbol{x}, \lambda) - \beta \right)_+ \right] \leq 0 \right\}, \quad (30)$$

where for each $i \in [I]$, $P_{q,i}(\boldsymbol{x}, \lambda)$ is defined in Proposition 5.

Proof According to the similar derivations in Proposition 3 and Proposition 5, we have

$$v_q^{\mathrm{CVaR}} = \min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \inf_{\beta \leq 0} \left\{ \varepsilon\beta + \lambda\theta^q + \mathbb{E}_{\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}} \left[\max_{i \in [I]} \left(\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) + P_{q,i}(\boldsymbol{x}, \lambda) - \beta \right)_+ \right] \right\} \leq 0, \lambda \geq 0 \right\}.$$

Note that the infimum is achievable since the left-hand function is continuous and convex in β and when $\beta \to -\infty$, the left-hand function goes to positive infinity. This completes the proof. □

Equivalently, we also recast the CVaR approximation (30) as a bilevel program. That is,

$$v_q^{\mathrm{CVaR}} = \min_t \quad t, \quad (31a)$$

$$\text{s.t.} \quad (\boldsymbol{x}^*, \lambda^*, \beta^*) \in \operatorname*{argmin}_{\substack{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{c}^\top \boldsymbol{x} \leq t, \\ \lambda \geq 0, \beta \leq 0}} \left\{ \varepsilon\beta + \lambda\theta^q + \mathbb{E}_{\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}} \left[\max_{i \in [I]} \left(\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) + P_{q,i}(\boldsymbol{x}, \lambda) - \beta \right)_+ \right] \right\}, \quad (31b)$$

$$\varepsilon\beta^* + \lambda^*\theta^q + \mathbb{E}_{\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}} \left[\max_{i \in [I]} \left(\boldsymbol{a}_i(\boldsymbol{x}^*)^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}^*) + P_{q,i}(\boldsymbol{x}^*, \lambda^*) - \beta \right)_+ \right] \leq 0. \quad (31c)$$

In the following result, we observe that under the same premise as Proposition 4, the CVaR approximation (30) can be simplified.

Proposition 7 *Suppose that* $\|\boldsymbol{a}_i(\boldsymbol{x})\|_* = \|\boldsymbol{a}_1(\boldsymbol{x})\|_*$ *for all* $i \in [I]$. *Then* CVaR *approximation* (30) *is*

$$v_q^{\text{CVaR}} = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \theta\varepsilon^{-\frac{1}{q}} \left\|\boldsymbol{a}_1(\boldsymbol{x})\right\|_* + \text{CVaR}_{1-\varepsilon}\left[\max_{i\in[I]}\left\{\boldsymbol{a}_i(\boldsymbol{x})^\top\tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x})\right\}\right] \leq 0 \right\}. \tag{32}$$

Proof Since $\|\boldsymbol{a}_i(\boldsymbol{x})\|_* = \|\boldsymbol{a}_1(\boldsymbol{x})\|_*$ for all $i \in [I]$, we must have $P_{q,i}(\boldsymbol{x},\lambda) = P_{q,1}(\boldsymbol{x},\lambda)$, for all $i \in [I]$. Then CVaR approximation is equivalent to

$$v_q^{\text{CVaR}} = \min_{\boldsymbol{x}\in\mathcal{X},\lambda,\beta} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \begin{array}{l} \lambda\theta^q + \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}}\left[\max\left\{\max_{i\in[I]}\left\{\boldsymbol{a}_i(\boldsymbol{x})^\top\tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) + P_{q,1}(\boldsymbol{x},\lambda)\right\},\beta\right\}\right] - (1-\varepsilon)\beta \leq 0, \\ \lambda \geq 0 \end{array} \right\}.$$

Subtracting the β in the inner maximum operator and redefining $\beta := \beta - P_{q,1}(\boldsymbol{x},\lambda)$, we have

$$v_q^{\text{CVaR}} = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \begin{array}{l} \lambda\theta^q + \varepsilon P_{q,1}(\boldsymbol{x},\lambda) + \varepsilon\beta + \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}}\left[\max_{i\in[I]}\left\{\boldsymbol{a}_i(\boldsymbol{x})^\top\tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) - \beta\right\}_+\right] \leq 0, \\ \lambda \geq 0 \end{array} \right\}.$$

Replacing the existence of β and $\lambda \geq 0$ by the minimum operator over β and $\lambda \geq 0$ in the left-hand side of the first constraint, we arrive at

$$v_q^{\text{CVaR}} = \min_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \colon \min_{\lambda\geq 0}\left\{\lambda\theta^q + \varepsilon P_{q,1}(\boldsymbol{x},\lambda)\right\} + \varepsilon\text{CVaR}_{1-\varepsilon}\left[\max_{i\in[I]}\left\{\boldsymbol{a}_i(\boldsymbol{x})^\top\tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x})\right\}\right] \leq 0 \right\}.$$

Note that for any given $\boldsymbol{x}$, the function $\lambda\theta^q + \varepsilon P_{q,1}(\boldsymbol{x},\lambda)$ is convex in λ over the domain $\lambda \in [0,\infty)$. Let us take its first-order derivative with respect to λ, and set it to be 0, which has a nonnegative root

$$\lambda^* = \varepsilon^{\frac{q-1}{q}} \left\|\boldsymbol{a}_1(\boldsymbol{x})\right\|_* q^{-1}\theta^{\frac{1}{q-1}} \geq 0.$$

Thus, λ^* solves $\min_{\lambda\geq 0}\left\{\lambda\theta^q + \varepsilon P_{q,1}(\boldsymbol{x},\lambda)\right\}$. Substituting λ^* into CVaR approximation, we arrive at the equivalent representation (30). □

B.4 Equivalent Reformulations of ALSO-X# and ALSO-X$\underline{\#}$

According to the reformulations above, under type $q-$Wasserstein ambiguity set with $q \in [1,\infty)$, ALSO-X# admits the form of

$$v_q^{A\#} = \min_t \quad t, \tag{33a}$$

$$\text{s.t.} \quad (\boldsymbol{x}^*,\lambda^*,\beta^*) \in \operatorname*{argmin}_{\substack{\boldsymbol{x}\in\mathcal{X},\boldsymbol{c}^\top\boldsymbol{x}\leq t,\\ \lambda\geq 0,\beta\leq 0}} \left\{ \varepsilon\beta + \lambda\theta^q + \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}}\left[\max_{i\in[I]}\left(\boldsymbol{a}_i(\boldsymbol{x})^\top\tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) + P_{q,i}(\boldsymbol{x},\lambda) - \beta\right)_+\right]\right\}, \tag{33b}$$

$$\boldsymbol{x}^* \in Z_q. \tag{33c}$$

Under type $q-$Wasserstein ambiguity set with $q \in [1,\infty)$, we show that ALSO-X# is better than CVaR approximation.

Proposition 1 *Under type* $q-$*Wasserstein ambiguity set with* $q \in [1,\infty)$, *suppose that for any objective upper bound* t *such that* $t \geq \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top\boldsymbol{x}$, ALSO-X# *is better than* CVaR *approximation.*

Proof For a given objective upper bound t, if the solution of lower-level CVaR approximation (31b) satisfies (31c), it is feasible to the DRCCP. Since the lower-level ALSO-X# (33b) and the lower-level CVaR approximation (31b) coincide, then ALSO-X# finds a feasible solution to the DRCCP if the CVaR approximation is able to find one. This completes the proof. □

Similarly, we introduce the ALSO-X$\underline{\#}$ by dropping the constraint $\beta \leq 0$ in the lower-level ALSO-X# (33b), which has the following formulation:

$$v_q^{A\underline{\#}} = \min_{t} \ t, \tag{34a}$$

$$\text{s.t.} \quad (\boldsymbol{x}^*, \lambda^*, \beta^*) \in \operatorname*{argmin}_{\substack{\boldsymbol{x}\in\mathcal{X}, \boldsymbol{c}^\top \boldsymbol{x} \leq t, \\ \lambda \geq 0, \beta}} \left\{ \varepsilon\beta + \lambda\theta^q + \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left[\max_{i\in[I]} \left(\boldsymbol{a}_i(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) + P_{q,i}(\boldsymbol{x}, \lambda) - \beta \right)_+ \right] \right\}, \tag{34b}$$

$$\boldsymbol{x}^* \in Z_q. \tag{34c}$$

Following the similar proof in Theorem 2, we can prove that ALSO-X# is better than ALSO-X$\underline{\#}$ under type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$.

Proposition 2 *Under type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$, suppose that for any objective upper bound t such that $t \geq \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x}$ and the lower-level* ALSO-X$\underline{\#}$ *admits a unique optimal $\boldsymbol{x}$-solution,* ALSO-X# *is better than* ALSO-X$\underline{\#}$.

Proof It is sufficient to show that for a given objective upper bound t, if the lower-level ALSO-X$\underline{\#}$ (34b) yields a feasible solution to DRCCP, i.e., satisfying (34c), then the lower-level ALSO-X# (33b) will also provide a feasible solution. Let $(\widehat{\boldsymbol{x}}, \widehat{\beta})$ denote an optimal solution from the lower-level ALSO-X# (33b) and let $(\bar{\boldsymbol{x}}, \bar{\beta})$ denote an optimal solution from the lower-level ALSO-X$\underline{\#}$ (34b). Suppose that $\bar{\boldsymbol{x}}$ is feasible to DRCCP (23), i.e., $\bar{\boldsymbol{x}} \in Z_q$ (24). Now let $\bar{\mathbb{P}}^*$ denote the worst-case distribution of $\tilde{\boldsymbol{\xi}}$ in the lower-level ALSO-X$\underline{\#}$ (34b) and let

$$\bar{\beta}^* := \bar{\mathbb{P}}^*\text{-VaR}_{1-\varepsilon} \left\{ \max_{i\in[I]} \boldsymbol{a}_i(\bar{\boldsymbol{x}})^\top \tilde{\boldsymbol{\xi}} - b_i(\bar{\boldsymbol{x}}) \right\} \leq \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-VaR}_{1-\varepsilon} \left\{ \max_{i\in[I]} \boldsymbol{a}_i(\bar{\boldsymbol{x}})^\top \tilde{\boldsymbol{\xi}} - b_i(\bar{\boldsymbol{x}}) \right\} \leq 0.$$

Then according to theorem 1 in [41] (see, e.g., equation (7) in [41]) and the discussions in Theorem 2, we have that $(\widehat{\boldsymbol{x}}, \bar{\beta}^*)$ is another optimal solution to the lower-level ALSO-X$\underline{\#}$ (34b). Since the only difference between the lower-level ALSO-X# (33b) and the lower-level ALSO-X$\underline{\#}$ (34b) is the constraint $\beta \leq 0$, with the assumption that the lower-level ALSO-X$\underline{\#}$ (34b) admits a unique optimal solution of $\boldsymbol{x}$, we must have $\bar{\boldsymbol{x}} = \widehat{\boldsymbol{x}}$. That is, for a given objective upper bound t, both lower-level problems have the same optimal value and optimal $\boldsymbol{x}$-solution. This implies that the lower-level ALSO-X# (33b) yields a feasible solution to DRCCP. □

We remark that the uniqueness of the lower-level ALSO-X$\underline{\#}$ can be achieved in many DRCCPs. For example, one condition is that the affine mappings are $\boldsymbol{a}_1(\boldsymbol{x}) = \boldsymbol{x}$, $b_1(\boldsymbol{x}) = b_1$, the random parameters $\tilde{\boldsymbol{\zeta}}$ follow a joint elliptical distribution with $\tilde{\boldsymbol{\zeta}} \sim \mathbb{P}_{\mathrm{E}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \widehat{g})$, and the norm defining the Wasserstein distance is the generalized Mahalanobis norm associated with the positive definite matrix $\boldsymbol{\Sigma}$. Under this setting, the lower-level ALSO-X$\underline{\#}$ can be written as

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}, \boldsymbol{c}^\top \boldsymbol{x} \leq t} \left\{ \overline{F}(\boldsymbol{x}) \colon = \boldsymbol{\mu}^\top \boldsymbol{x} + \left[\overline{G} \left(\left(\Phi^{-1}(1-\varepsilon) \right)^2 / 2 \right) / \varepsilon + \theta\varepsilon^{-\frac{1}{q}} \right] \sqrt{\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}} - b_1 \right\},$$

and its first-order and second-order derivatives are

$$\frac{\partial \overline{F}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{\mu}^\top + \left[\overline{G} \left(\left(\Phi^{-1}(1-\varepsilon) \right)^2 / 2 \right) / \varepsilon + \theta\varepsilon^{-\frac{1}{q}} \right] \boldsymbol{\Sigma} \boldsymbol{x}^\top,$$

$$\frac{\partial^2 \overline{F}(\boldsymbol{x})}{\partial \boldsymbol{x}^2} = \left[\overline{G} \left(\left(\Phi^{-1}(1-\varepsilon) \right)^2 / 2 \right) / \varepsilon + \theta\varepsilon^{-\frac{1}{q}} \right] \boldsymbol{\Sigma} \succ \boldsymbol{0}.$$

Hence, the lower-level ALSO-X$\underline{\#}$ admits a unique solution whenever set $\mathcal{X}$ is convex.

We conclude this section by providing theoretical comparisons among the output objective values of ALSO-X#, ALSO-X$\underline{\#}$, ALSO-X, and CVaR approximation under type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$, which are shown in Figure 3.

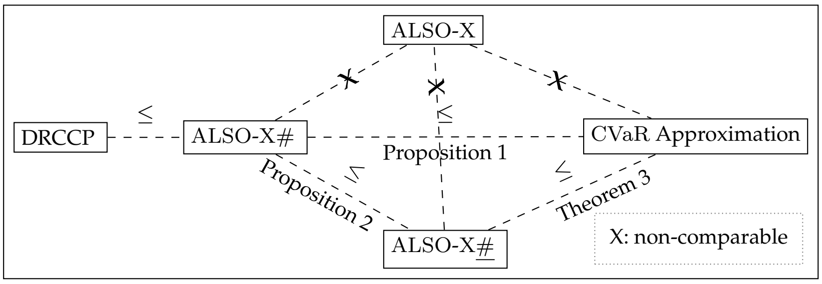


Fig. 3: Summary of Comparisons under Type $q-$Wasserstein ambiguity set with $q \in [1, \infty)$