On Distributionally Robust Chance Constrained Program with Wasserstein Distance

Weijun Xie*[1]

[1]Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061

June 21, 2018

Abstract

This paper studies a distributionally robust chance constrained program (DRCCP) with Wasserstein ambiguity set, where the uncertain constraints should satisfy with a probability at least a given threshold for all the probability distributions of the uncertain parameters within a chosen Wasserstein distance from an empirical distribution. In this work, we investigate equivalent reformulations and approximations of such problems. We first show that a DRCCP can be reformulated as a conditional-value-at-risk constrained optimization problem, and thus admits tight inner and outer approximations. When the metric space of uncertain parameters is a normed vector space, we show that a DRCCP of bounded feasible region is mixed integer representable by introducing big-M coefficients and additional binary variables. For a DRCCP with pure binary decision variables, by exploring submodular structure, we show that it admits a big-M free formulation and can be solved by branch and cut algorithm. This result can be generalized to mixed integer DRCCPs. Finally, we present a numerical study to illustrate effectiveness of the proposed methods.

*Email: wxie@vt.edu.

1 Introduction

1.1 Setting

A distributionally robust chance constrained program (DRCCP) is of the form:

$$\min\ \boldsymbol{c}^\top \boldsymbol{x}, \tag{1a}$$

$$\text{s.t. } \boldsymbol{x} \in S, \tag{1b}$$

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : \boldsymbol{a}(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}}_i \le b_i(\boldsymbol{x}), \forall i \in [I]\right\} \ge 1-\epsilon. \tag{1c}$$

In (1), the vector $\boldsymbol{x} \in \mathbb{R}^n$ denotes the decision variables; the vector $\boldsymbol{c} \in \mathbb{R}^n$ denotes the objective function coefficients; the set $S \subseteq \mathbb{R}^n$ denotes deterministic constraints on $\boldsymbol{x}$; and the constraint (1c) is a chance constraint involving I uncertain inequalities specified by the random vectors $\tilde{\boldsymbol{\xi}}_i$ supported on set $\Xi_i \subseteq \mathbb{R}^n$ for each $i \in [I]$ with a joint probability distribution $\mathbb{P}$ from a family $\mathcal{P}$, termed "*ambiguity set*". We let $[R] := \{1, 2, \ldots, R\}$ for any positive integer R, and for each uncertain constraint $i \in [I]$, $\boldsymbol{a}(\boldsymbol{x}) \in \mathbb{R}^n$ and $b_i(\boldsymbol{x}) \in \mathbb{R}$ denote affine mappings of $\boldsymbol{x}$ such that $\boldsymbol{a}(\boldsymbol{x}) = \eta\boldsymbol{x} + (1-\eta)\boldsymbol{e}$ and $b_i(\boldsymbol{x}) = \boldsymbol{B}_i^\top\boldsymbol{x} + b^i$ with $\eta \in \{0, 1\}$, all-one vector $\boldsymbol{e} \in \mathbb{R}^n$, $\boldsymbol{B}_i \in \mathbb{R}^n$, and $b^i \in \mathbb{R}$, respectively. For notational convenience, we let $\Xi = \prod_{i\in[I]} \Xi_i$ and $\tilde{\boldsymbol{\xi}} = (\tilde{\boldsymbol{\xi}}_1, \ldots, \tilde{\boldsymbol{\xi}}_I)$. Note that (i) for any $i, j \in [I]$ and $i \neq j$, random vectors $\tilde{\boldsymbol{\xi}}_i$ and $\tilde{\boldsymbol{\xi}}_j$ can be correlated; and (ii) we use $\eta \in \{0, 1\}$ to differentiate whether (1c) involves left-hand uncertainty (i.e., $\eta = 1$) or right-hand uncertainty (i.e., $\eta = 0$).

The chance constraint (1c) requires that all I uncertain constraints are simultaneously satisfied for all the probability distributions from ambiguity set $\mathcal{P}$ with a probability at least $(1-\epsilon)$, where $\epsilon \in (0, 1)$ is a specified risk tolerance. We call (1) a *single* DRCCP if $I = 1$ and a *joint* DRCCP if $I \ge 2$. Also, (1) is termed a DRCCP with *right-hand* uncertainty if $\eta = 0$ and a DRCCP with *left-hand* uncertainty, otherwise. For a joint DRCCP, if $I = 2, \tilde{\boldsymbol{\xi}}_1 = -\tilde{\boldsymbol{\xi}}_2$, we call (1) as a *two-sided* DRCCP.

We denote the feasible region induced by (1c) as

$$Z := \left\{\boldsymbol{x} \in \mathbb{R}^n : \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : \boldsymbol{a}(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}}_i \le b_i(\boldsymbol{x}), \forall i \in [I]\right\} \ge 1-\epsilon\right\}. \tag{2}$$

In this paper, we consider Wassertein ambiguity set $\mathcal{P}$.

(A1) The Wasserstein ambiguity set $\mathcal{P}$ is defined as

$$\mathcal{P} = \left\{\mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}\times\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}}[d(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\zeta}})] \le \delta\right\}, \tag{3}$$

where $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ denotes the discrete empirical distribution of $\tilde{\boldsymbol{\zeta}}$ on the countable support $\mathcal{Z} = \{\boldsymbol{\zeta}^j\}_{j\in[N]} \subseteq \Xi$ with point mass function $\{p_j\}_{j\in[N]}$, $d : \Xi \times \Xi \to \mathbb{R}_+$ denotes distance metric and $\delta > 0$ denotes the Wasserstein radius. We assume that (Ξ, d) is a totally bounded Polish (separable complete metric) space with distance metric d, i.e., for every $\widehat{\epsilon} > 0$, there exists a finite covering of Ξ by balls with radius at most $\widehat{\epsilon}$.

Note that The Wasserstein metric measures the distance between true distribution and empirical distribution and is able to recover the true distribution when the number of sampled data goes to infinity [14].

1.2 Related Literature

There are significant work on reformulation, convexity and approximations of set Z under various ambiguity sets [7, 18, 19, 21, 37, 39]). For a single DRCCP, when $\mathcal{P}$ consists of all probability distributions with given first and second moments, the set Z is second-order conic representable [7, 13]. Similar convexity results hold for single DRCCP when $\mathcal{P}$ also incorporates other distributional information such as the support of $\tilde{\boldsymbol{\xi}}$ [10], the unimodality of $\mathbb{P}$ [18, 24], or arbitrary convex mapping of $\tilde{\boldsymbol{\xi}}$ [37]. For a joint DRCCP, [19] provided the first convex reformulation of Z in the absence of coefficient uncertainty, i.e. $\eta = 0$, when $\mathcal{P}$ is characterized by the mean, a positively homogeneous dispersion measure, and a conic support of $\tilde{\boldsymbol{\xi}}$. For the more general coefficient uncertainty setting, [37] identified several sufficient conditions for Z to be convex (e.g., when $\mathcal{P}$ is specified by one moment constraint), and [36] showed that Z is convex for two-sided DRCCP when $\mathcal{P}$ is characterized by the first two moments.

When DRCCP set Z is not convex, many inner convex approximations have been proposed. In [9], the authors proposed to aggregate the multiple uncertain constraints with positive scalars in to a single constraint, and then use conditional-value-at-risk (CVaR) approximation scheme [28] to develop an inner approximation of Z. This approximation is shown to be exact for single DRCCP when $\mathcal{P}$ is specified by first and second moments in [43] or, more generally, by convex moment constraints in [37]. In [38], the authors provided several sufficient conditions under which the well-known Bonferroni approximation of joint DRCCP is exact and yields a convex reformulation.

Recently, there are many successful developments on data driven distributionally robust programs with Wasserstein ambiguity set (3) [16, 27, 41]. For instance, [16, 27] studied its reformulation under different settings. Later on, [4, 15, 23, 32] applied it to the optimization problems related with machine learning. Other relevant works can be found [3, 17, 22, 26]. However, there is very limited literature on DRCCP with Wasserstein ambiguity set. In [35], the authors proved that it is strongly NP-hard to optimize over the DRCCP set Z with Wasserstein ambiguity set and proposed a bicriteria approximation for a class of DRCCP with covering uncertain constraints (i.e., S is a closed convex cone and $\Xi_i \in \mathbb{R}^n_-, \boldsymbol{B}_i \in \mathbb{R}^n_+, b_i \in \mathbb{R}_-$ for each $i \in [I]$). In [11], the authors considered two-sided DRCCP with right-hand uncertainty and proposed its tractable reformulation, while in [20], the authors studied CVaR approximation of DRCCP. As far as the author is concerned, there is no work on developing tight approximations and exact reformulations of general DRCCP with Wasserstein ambiguity set.

1.3 Contributions

In this paper, we study approximations and exact reformulations of DRCCP under Wasserstein ambiguity set. In particular, our main contributions are summarized as below.

1. We derive a deterministic equivalent reformulation for set Z and show that this reformulation admits a conditional-value-at-risk (CVaR) interpretation. Based upon this fact, we are able to derive tight inner and outer approximations.

2. When the support Ξ is an $n \times I$- dimensional vectlor space and the distance metric is a norm (i.e, $d(\boldsymbol{\xi}, \boldsymbol{\zeta}) = \|\boldsymbol{\xi} - \boldsymbol{\zeta}\|$), we show that the feasible region $S \cap Z$ of a DRCCP, once bounded, is mixed integer representable with big-M coefficients and $N \times I$ additional binary variables. We also derive compact formulations for the proposed inner and outer approximations and compare their strengths.

3. When the decision variables are pure binary (i.e., $S \subseteq \{0,1\}^n$), we first show that the nonlinear constraints in the reformulation can be recast as submodular knapsack constraints. Then, by exploring the polyhedral properties of submodular functions, we propose a new big-M free mixed integer linear reformulation. In a numerical study, we further show that the proposed formulation can be effectively solved by branch and cut algorithm.

The remainder of the paper is organized as follows. Section 2 presents exact reformulation of DRCCP set Z as well as its inner and outer approximations under a general setting. Section 3 provides (mixed integer) convex reformulations of feasible region $S \cap Z$ and its inner and outer approximations when the metric space of the random variables is a normed vector space. Section 4 studies binary DRCCP (i.e., $S \subseteq \{0,1\}^n$), develops a big-M free formulation, and numerically illustrates the proposed methods. Section 5 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 or matrices, and use corresponding non-bold letters to denote their components. We let $\boldsymbol{e}$ be the all-ones vector, and let $\boldsymbol{e}_i$ be the ith standard basis vector. Given an integer n, we let $[n] := \{1, 2, \ldots, n\}$, and use $\mathbb{R}^n_+ := \{\boldsymbol{x} \in \mathbb{R}^n : x_l \geq 0, \forall l \in [n]\}$ and $\mathbb{R}^n_- := \{\boldsymbol{x} \in \mathbb{R}^n : x_l \leq 0, \forall l \in [n]\}$. Given a real number t, we let $(t)_+ := \max\{t, 0\}$. Given a finite set I, we let $|I|$ denote its cardinality. We let $\tilde{\boldsymbol{\xi}}$ denote a random vector with support Ξ and denote one of its realization by $\boldsymbol{\xi}$. Given a set R, the characteristic function $\chi_R(\boldsymbol{x}) = 0$ if $\boldsymbol{x} \in R$, and ∞, otherwise, while the indicator function $\mathbb{I}(\boldsymbol{x} \in R)$ =1 if $\boldsymbol{x} \in R$, and 0, otherwise. For a matrix $\boldsymbol{A}$, we let $\boldsymbol{A}_{i\bullet}$ denote ith row of $\boldsymbol{A}$ and $\boldsymbol{A}_{\bullet j}$ denote jth column of $\boldsymbol{A}$. Additional notation will be introduced as needed. Given a subset $T \subseteq [n]$, we define an n-dimensional binary vector $\boldsymbol{e}_T$ as $(\boldsymbol{e}_T)_\tau = \begin{cases} 1, & \text{if } \tau \in T \\ 0, & \text{if } \tau \in [n] \setminus T \end{cases}$.

2 General Case: Reformulations and Approximations

In this section, we will study an equivalent reformulation of set Z under Assumption (A1). This reformulation has conditional-value-at-risk (CVaR) interpretation, therefore allows us to derive tight inner and outer approximations.

2.1 Exact Reformulation

In this part, we will reformulate set Z into its deterministic counterpart. The main idea of this reformulation is that we first use the strong duality result from [16] to formulate the worst-case chance constraint into its dual form, and then break down the indicator function according to its definition.

Theorem 1. *Under Assumption (A1), set Z is equivalent to*

$$Z = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \delta\lambda - \epsilon \leq \frac{1}{N} \sum_{j \in [N]} \min \left\{ \lambda f(\boldsymbol{x}, \boldsymbol{\zeta}^j) - 1, 0 \right\}, \lambda \geq 0 \right\}, \tag{4}$$

where

$$f(\boldsymbol{x}, \boldsymbol{\zeta}) := \min_{i \in [I]} \inf_{\boldsymbol{\xi} \in \Xi, \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x})} d(\boldsymbol{\xi}, \boldsymbol{\zeta}). \tag{5}$$

Proof. Note that

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : \boldsymbol{a}(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}}_i \le b_i(\boldsymbol{x}), \forall i \in [I]\right\} \ge 1-\epsilon$$

is equivalent to

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : \boldsymbol{a}(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}}_i > b_i(\boldsymbol{x}), \exists i \in [I]\right\} \le \epsilon.$$

By Theorem 1 in [16], $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} : \boldsymbol{a}(\boldsymbol{x})^\top \tilde{\boldsymbol{\xi}}_i > b_i(\boldsymbol{x}), \exists i \in [I]\right\}$ is equivalent to

$$\min_{\lambda\ge 0} \lambda\delta - \frac{1}{N}\sum_{j\in[N]} \inf_{\boldsymbol{\xi}\in\Xi} \left[\lambda d(\boldsymbol{\xi},\boldsymbol{\zeta}^j) - \mathbb{I}\left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x}), \exists i \in [I]\right)\right]. \tag{6a}$$

Thus set Z is equivalent to

$$Z := \left\{\boldsymbol{x}\in\mathbb{R}^n : \lambda\delta - \frac{1}{N}\sum_{j\in[N]} \inf_{\boldsymbol{\xi}\in\Xi} \left[\lambda d(\boldsymbol{\xi},\boldsymbol{\zeta}^j) - \mathbb{I}\left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x}), \exists i \in [I]\right)\right] \le \epsilon, \lambda \ge 0\right\}. \tag{6b}$$

We now break down the indicator function in the infimum of (6b) and reformulate it as below.

Claim 1. *for given $\lambda \ge 0$ and $\boldsymbol{\zeta} \in \mathcal{Z}$, we have*

$$\inf_{\boldsymbol{\xi}\in\Xi} \left[\lambda d(\boldsymbol{\xi},\boldsymbol{\zeta}) - \mathbb{I}\left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x}), \exists i \in [I]\right)\right] = \min\left\{\min_{i\in[I]} \inf_{\boldsymbol{\xi}\in\Xi, \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x})} \left[\lambda d(\boldsymbol{\xi},\boldsymbol{\zeta}) - 1\right], 0\right\}. \tag{6c}$$

Proof. We first note that $\mathbb{I}\left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x}), \exists i \in [I]\right) = \max_{i\in[I]} \mathbb{I}\left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x})\right)$. Thus,

$$\inf_{\boldsymbol{\xi}\in\Xi} \left[\lambda d(\boldsymbol{\xi},\boldsymbol{\zeta}) - \mathbb{I}\left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x}), \exists i \in [I]\right)\right] = \min_{i\in[I]} \inf_{\boldsymbol{\xi}\in\Xi} \left[\lambda d(\boldsymbol{\xi},\boldsymbol{\zeta}) - \mathbb{I}\left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x})\right)\right].$$

Therefore, we only need to show that for any $i \in [I]$,

$$\inf_{\boldsymbol{\xi}\in\Xi} \left[\lambda d(\boldsymbol{\xi},\boldsymbol{\zeta}) - \mathbb{I}\left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x})\right)\right] = \min\left\{\inf_{\boldsymbol{\xi}\in\Xi, \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x})} \left[\lambda d(\boldsymbol{\xi},\boldsymbol{\zeta}) - 1\right], 0\right\}. \tag{6d}$$

There are two cases:

Case 1. If $\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i > b_i(\boldsymbol{x})$, then in the left-hand side of (6d), the infimum is equal to -1 by letting $\boldsymbol{\xi} := \boldsymbol{\zeta}$, which equals to the right-hand side since the infimum is also achieved by $\boldsymbol{\xi} := \boldsymbol{\zeta}$.

Case 2. If $\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i \le b_i(\boldsymbol{x})$, then for any $\boldsymbol{\xi} \in \Xi$, we either have $\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x})$ or $\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i \le b_i(\boldsymbol{x})$. Hence, the left-hand side of (6d) is equivalent to

$$\begin{aligned}
&\inf_{\boldsymbol{\xi}\in\Xi} \left[\lambda d(\boldsymbol{\xi},\boldsymbol{\zeta}) - \mathbb{I}\left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x})\right)\right] \\
&= \min\left\{\inf_{\boldsymbol{\xi}\in\Xi, \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x})} \left[\lambda d(\boldsymbol{\xi},\boldsymbol{\zeta}) - 1\right], \inf_{\boldsymbol{\xi}\in\Xi, \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i \le b_i(\boldsymbol{x})} \left[\lambda d(\boldsymbol{\xi},\boldsymbol{\zeta})\right]\right\} \\
&= \min\left\{\inf_{\boldsymbol{\xi}\in\Xi, \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i > b_i(\boldsymbol{x})} \left[\lambda d(\boldsymbol{\xi},\boldsymbol{\zeta}) - 1\right], 0\right\},
\end{aligned}$$

where $\inf_{\boldsymbol{\xi}\in\Xi, \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i \le b_i(\boldsymbol{x})} \left[d(\boldsymbol{\xi},\boldsymbol{\zeta})\right] = 0$ by letting $\boldsymbol{\xi} := \boldsymbol{\zeta}$.

$\diamond$

Thus, By Claim 1, set Z is equivalent to (4).$\square$

In Theorem 1, we must have $\lambda > 0$, thus can define a new variable $\gamma = \frac{1}{\lambda}$ in (4) and reformulate set Z into the following equivalent form.

Theorem 2. *Under Assumption (A1), set Z is equivalent to*

$$Z = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \delta - \epsilon\gamma \leq \frac{1}{N} \sum_{j \in [N]} \min\left\{ f(\boldsymbol{x}, \boldsymbol{\zeta}^j) - \gamma, 0 \right\}, \gamma \geq 0 \right\}, \tag{7}$$

where function $f(\cdot, \cdot)$ is defined in (5).

Proof. Next, let Z' denote the set in the right-hand side of (7), we only need to show that sets Z, Z' are equivalent, i.e., $Z = Z'$.

$(Z \subseteq Z')$ Given $\boldsymbol{x} \in Z$, there exists $\lambda \geq 0$ such that $(\boldsymbol{x}, \lambda)$ satisfies (4). If $\lambda > 0$, then let $\gamma = \frac{1}{\lambda}$. Then it is easy to see that $(\boldsymbol{x}, \gamma)$ satisfies (7). Hence, $\boldsymbol{x} \in Z'$.

Now suppose that $\lambda = 0$, then in (4), we have

$$-\epsilon \leq -1$$

a contradiction that $\epsilon \in (0, 1)$.

$(Z \supseteq Z')$ Similarly, given $\boldsymbol{x} \in Z'$, there exists $\gamma \geq 0$ such that $(\boldsymbol{x}, \gamma)$ satisfies (7). If $\gamma > 0$, then let $\lambda = \frac{1}{\gamma}$. Then it is easy to see that $(\boldsymbol{x}, \lambda)$ satisfies (4). Hence, $\boldsymbol{x} \in Z$.

Now suppose that $\gamma = 0$, then in (7), we have

$$\min\left\{ f(\boldsymbol{x}, \boldsymbol{\zeta}^j) - \gamma, 0 \right\} := 0$$

for each $j \in [N]$. Thus, (7) reduces to $\delta \leq 0$ contradicting that $\delta > 0$.

$\square$

Before showing a conditional-value-at-risk (CVaR) interpretation of set Z, let us begin with the following two definitions. Given a random variable $\tilde{X}$, let $\mathbb{P}$ and $F_{\tilde{X}}(\cdot)$ be its probability distribution and cumulative distribution function, respectively. Then $(1-\epsilon)$-value at risk (VaR) of $\tilde{X}$ is

$$\mathbf{VaR}_{1-\epsilon}(\tilde{X}) := \min\left\{ s : F_{\tilde{X}}(s) \geq 1 - \epsilon \right\},$$

while its $(1-\epsilon)$-conditional-value-at-risk (CVaR) [31] is defined as

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

With the definitions above, we observe that set Z in (7) has a CVaR interpretation.

Corollary 1. *Under Assumption (A1), set Z is equivalent to*

$$Z = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \frac{\delta}{\epsilon} + \mathbf{CVaR}_{1-\epsilon}\left[-f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}})\right] \leq 0 \right\}, \tag{8}$$

where $f(\cdot,\cdot)$ is defined in (5), *and* $\mathbf{CVaR}_{1-\epsilon}\left[-f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}})\right] = \min_\gamma \left\{ \gamma + \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}}\left[-f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}}) - \gamma\right]_+ \right\}$.

Proof. First, we observe that the constraint in (7) directly implies $\gamma > 0$, thus the nonnegativity constraint of γ can be dropped, i.e., equivalently, we have

$$Z = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \frac{\delta}{\epsilon} - \gamma + \frac{1}{N\epsilon} \sum_{j \in [N]} \max\left\{-f(\boldsymbol{x}, \boldsymbol{\zeta}^j) + \gamma, 0\right\} \leq 0 \right\}.$$

Next, in the above formulation, let $\gamma := -\gamma$ and replace the existence of γ by finding the best γ such that the constraint still holds, we arrive at

$$Z = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \frac{\delta}{\epsilon} + \min_\gamma \left\{ \gamma + \frac{1}{N\epsilon} \sum_{j \in [N]} \max\left\{-f(\boldsymbol{x}, \boldsymbol{\zeta}^j) - \gamma, 0\right\} \right\} \leq 0 \right\},$$

which is equivalent to (8).□

In the following sections, our derivations of exact reformulations is based upon Theorem 2 while the approximations are mainly according to $\mathbf{CVaR}$ interpretation in Corollary 1.

2.2 Outer and Inner Approximations

In this subsection, we will introduce one outer approximation and three different inner approximations by exploring the exact reformulations in the previous section.

VaR Outer Approximation: Note from [31] that for any random variable $\tilde{X}$, we have

$$\mathbf{CVaR}_{1-\epsilon}\left(\tilde{X}\right) = \mathbf{VaR}_{1-\epsilon}\left(\tilde{X}\right) + \frac{1}{\epsilon}\mathbb{E}\left[\tilde{X} - \mathbf{VaR}_{1-\epsilon}\left(\tilde{X}\right)\right]_+ \geq \mathbf{VaR}_{1-\epsilon}\left(\tilde{X}\right).$$

Therefore, in Corollary 1, if we replace $\mathbf{CVaR}_{1-\epsilon}(\cdot)$ by $\mathbf{VaR}_{1-\epsilon}(\cdot)$, then we have the following outer approximation of set Z.

Theorem 3. *Under Assumption (A1), set Z is outer approximated by*

$$Z_{\mathbf{VaR}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \mathbb{P}_{\tilde{\boldsymbol{\zeta}}}\left\{ f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}}) \geq \frac{\delta}{\epsilon} \right\} \geq 1 - \epsilon \right\}. \tag{9}$$

Proof. Due to the well known result in [31] that $\mathbf{CVaR}_{1-\epsilon}\left[-f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}})\right] \geq \mathbf{VaR}_{1-\epsilon}\left[-f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}})\right]$. Therefore, set Z is outer approximated by

$$Z_{\mathbf{VaR}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \frac{\delta}{\epsilon} + \mathbf{VaR}_{1-\epsilon}\left[-f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}})\right] \leq 0 \right\},$$

which is equivalent to

$$Z_{\mathbf{VaR}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \mathbb{P}_{\tilde{\zeta}} \left\{ f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}}) \geq \frac{\delta}{\epsilon} \right\} \geq 1 - \epsilon \right\}.$$

Inner Approximation I- Robust Scenario Approximation: On the other hand, we notice that for any random variable $\tilde{X}$, we have

$$\mathbf{CVaR}_{1-\epsilon} \left(\tilde{X} \right) \leq \mathbf{CVaR}_1 \left(\tilde{X} \right) := \text{ess.}\sup(\tilde{X}).$$

Thus, in Corollary 1, if we replace $\mathbf{CVaR}_{1-\epsilon}(\cdot)$ by $\text{ess.}\sup(\cdot)$, then we have the following inner approximation of set Z.

Theorem 4. *Under Assumption (A1), set Z is inner approximated by*

$$Z_S = \left\{ \boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}, \boldsymbol{\zeta}^j) \geq \frac{\delta}{\epsilon}, \forall j \in [N] \right\}. \tag{10}$$

Proof. Since $\mathbf{CVaR}_{1-\epsilon} \left[-f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}}) \right] \leq \text{ess.}\sup \left[-f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}}) \right]$, and $\tilde{\boldsymbol{\zeta}}$ is a discrete random vector, therefore, set Z can inner approximated by

$$Z_S = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \mathbb{P}_{\tilde{\zeta}} \left\{ f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}}) \geq \frac{\delta}{\epsilon} \right\} = 1 \right\},$$

which is equivalent to (10). □

Note that set Z_S has a similar structure as scenario approach to chance constrained problem [6, 8, 29], and indeed can be viewed as a "robust" scenario approach to chance constrained problem. We will discuss more on this fact in Section 3.2.

Inner Approximation II- Inner Chance Constrained Approximation: Next we propose a chance constrained inner approximation of DRCCP set Z by constructing a feasible γ in (7).

Theorem 5. *Under Assumption (A1) and $\epsilon \in (0,1)$, $N\epsilon \notin \mathbb{Z}_+$, set Z is inner approximated by*

$$Z_I = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \mathbb{P}_{\tilde{\zeta}} \left\{ f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}}) \geq \frac{\delta}{\epsilon - \alpha} \right\} \geq 1 - \alpha, 0 \leq \alpha \leq \frac{\lfloor N\epsilon \rfloor}{N} \right\}. \tag{11}$$

Proof. For any $\boldsymbol{x} \in Z_I$, we would like to show that $\boldsymbol{x} \in Z$. Since $\boldsymbol{x} \in Z_I$, there exists an α such that $(\boldsymbol{x}, \alpha)$ satisfies constraints in (11). Now let us define $\gamma = \frac{N\delta}{N\epsilon - \lfloor N\alpha \rfloor}$. Let define a set

$$\mathcal{C} = \left\{ j \in [N] : f(\boldsymbol{x}, \boldsymbol{\zeta}^j) \leq \gamma \right\}.$$

Since $\frac{\delta}{\epsilon - \alpha} \geq \frac{N\delta}{N\epsilon - \lfloor N\alpha \rfloor} := \gamma$, thus by (11), $|\mathcal{C}| \leq \lfloor N\alpha \rfloor$. Hence,

$$\frac{1}{N} \sum_{j \in [N]} \min \left\{ f(\boldsymbol{x}, \boldsymbol{\zeta}^j) - \gamma, 0 \right\} = \frac{1}{N} \sum_{j \in \mathcal{C}} \left(f(\boldsymbol{x}, \boldsymbol{\zeta}^j) - \gamma \right) \geq -\frac{|\mathcal{C}|}{N} \gamma \geq -\frac{\lfloor N\alpha \rfloor}{N} \gamma = \delta - \epsilon\gamma,$$

where the first inequality is due to $f(\boldsymbol{x}, \boldsymbol{\zeta}^j) \geq 0$ and the second inequality is due to $|\mathcal{C}| \leq \lfloor N\alpha \rfloor$.□

We remark that this result together with set $Z_{\mathbf{VaR}}$ shows that DRCCP set Z can be inner and outer approximated by regular chance constraints with empirical probability distribution $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$.

We also note that (i) set Z_S is a special case of set Z_I by letting $\alpha = 0$, thus, we must have $Z_S \subseteq Z_I$; (ii) there are $\lfloor N\epsilon \rfloor + 1$ non-dominant α values, that is, we must have $\alpha \in \{0, \frac{1}{N}, \ldots, \frac{\lfloor N\epsilon \rfloor}{N}\}$. Indeed, suppose that $\alpha \in (\frac{i-1}{N}, \frac{i}{N})$ for an $i \in [\lfloor N\epsilon \rfloor]$, then the feasible region expands if we decrease the value of α to $\frac{i-1}{N}$. Therefore, to optimize over set Z_I, we can enumerate these $\lfloor N\epsilon \rfloor + 1$ values of α and choose the one which yields the smallest objective value. These two results are summarized below.

Corollary 2. *Suppose that Assumption (A1) holds and $\epsilon \in (0,1)$, $N\epsilon \notin \mathbb{Z}_+$ and set Z_I is defined in* (11), *then*

(i) $Z_S \subseteq Z_I$; *and*

(ii) set Z_I *is equivalent to*

$$Z_I = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \mathbb{P}_{\tilde{\boldsymbol{\zeta}}} \left\{ f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}}) \geq \frac{\delta}{\epsilon - \alpha} \right\} \geq 1 - \alpha, \alpha \in \left\{ 0, \frac{1}{N}, \ldots, \frac{\lfloor N\epsilon \rfloor}{N} \right\} \right\}. \tag{12}$$

Inner Approximation III- CVaR Approximation: Finally, we close this section by studying a well known convex approximation of a chance constraint, which is to replace the nonconvex chance constraint by a convex constraint defined by CVaR (cf., [28]). For a DRCCP, the resulting formulation is

$$Z_{\mathbf{CVaR}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \sup_{\mathbb{P} \in \mathcal{P}} \inf_{\beta} \left[-\epsilon\beta + \mathbb{E}_{\mathbb{P}} \left(\max_{i \in [I]} \Big(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi} - b_i(\boldsymbol{x}) \Big) + \beta \right)_+ \right] \leq 0. \right\} \tag{13}$$

Set $Z_{\mathbf{CVaR}}$ (13) is convex and is an inner approximation of set Z. The following results show a reformulation of set $Z_{\mathbf{CVaR}}$. We would like to acknowledge that this result has been independently observed by a recent work in [20]. For the completeness of this paper, we present a proof with our notation as below.

Theorem 6. *Set $Z_{\mathbf{CVaR}} \subseteq Z$ and is equivalent to*

$$Z_{\mathbf{CVaR}} = \left\{ x : \begin{array}{l} -\epsilon\beta + \lambda\delta - \dfrac{1}{N} \displaystyle\sum_{j \in [N]} \inf_{\boldsymbol{\xi} \in \Xi} \left[\lambda d(\boldsymbol{\xi}, \boldsymbol{\zeta}^j) - \left(\max_{i \in [I]} \Big(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i - b_i(\boldsymbol{x}) \Big) + \beta \right)_+ \right] \leq 0, \\ \lambda, \beta \geq 0. \end{array} \right\}$$

(14a)

(14b)

Proof. Note that Wassersetein ambiguity set $\mathcal{P}$ is weakly compact [5], thus according to Theorem 2.1 in [33], set $Z_{\mathbf{CVaR}}$ is equivalent to

$$Z_{\mathbf{CVaR}} = \left\{ x : \inf_{\beta} \left[-\epsilon\beta + \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left(\max_{i \in [I]} \Big(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi} - b_i(\boldsymbol{x}) \Big) + \beta \right)_+ \right] \leq 0. \right\} \tag{15}$$

Note that in the (15), the infimum must be achieved. Indeed, we first note that for any $\beta < 0$, the inequality in (15) will not be satisfied. Thus, we must have $\beta \geq 0$. On the other hand, we note that

$$-\epsilon\beta + \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \left(\max_{i\in[I]} \left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi} - b_i(\boldsymbol{x}) \right) + \beta \right)_+ \geq -\epsilon\beta + \mathbb{E}_{\mathbb{P}_{\tilde{\zeta}}} \left(\max_{i\in[I]} \left(\boldsymbol{a}(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}} - b_i(\boldsymbol{x}) \right) + \beta \right)_+$$

where the inequality is due to $\mathbb{P}_{\tilde{\zeta}} \in \mathcal{P}$. The right-hand side of the above inequality will be equal to $(1-\epsilon)\beta > 0$ for any $\beta > \left(\max_{i\in[I], j\in[N]} \left(b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j \right), 0 \right)$. Thus, the best β in (15) is bounded, i.e., $Z_{\mathbf{CVaR}}$ is equivalent to

$$Z_{\mathbf{CVaR}} = \left\{ x : -\epsilon\beta + \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \left(\max_{i\in[I]} \left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i - b_i(\boldsymbol{x}) \right) + \beta \right)_+ \leq 0, \beta \geq 0. \right\} \tag{16}$$

By Theorem 1 in [16], the above formulation is further equal to

$$Z_{\mathbf{CVaR}} = \left\{ \begin{array}{l} x : \min_{\lambda \geq 0} \left[-\epsilon\beta + \lambda\delta - \dfrac{1}{N} \displaystyle\sum_{j\in[N]} \inf_{\boldsymbol{\xi}\in\Xi} \left[\lambda d(\boldsymbol{\xi}, \boldsymbol{\zeta}^j) - \left(\max_{i\in[I]} \left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi} - b_i(\boldsymbol{x}) \right) + \beta \right)_+ \right] \right] \leq 0, \\ \beta \geq 0, \end{array} \right\}$$

which is equivalent to (14).□

3 DRCCP with Normed Vector Space (Ξ, d)

Note that the results in the previous section are quite general. In this section, we will show that these results can be significantly simplified given that (Ξ, d) is a normed vector space. In particular, we make the following assumption.

(A2) The support Ξ is an $n \times I$-dimensional vector space and distance metric $d(\boldsymbol{\xi}, \boldsymbol{\zeta}) = \|\boldsymbol{\xi} - \boldsymbol{\zeta}\|$.

3.1 Exact Mixed Integer Program Reformulation

In this subsection, we show that set Z is mixed integer representable under Assumptions (A1)-(A2). To begin with, we observe that under additional Assumption (A2), function $f(\cdot, \cdot)$ in Theorem 2 can be explicitly calculated. Thus,

Theorem 7. *Under Assumptions (A1)-(A2), set Z is equivalent to*

$$Z = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \delta - \epsilon\gamma \leq \dfrac{1}{N} \displaystyle\sum_{j\in[N]} \min\left\{ f(\boldsymbol{x}, \boldsymbol{\zeta}^j) - \gamma, 0 \right\}, & \text{(17a)} \\ \gamma \geq 0, & \text{(17b)} \end{array} \right\}$$

where

$$f(\boldsymbol{x}, \boldsymbol{\zeta}) = \min\left\{ \min_{i\in[I]\setminus\mathcal{I}(\boldsymbol{x})} \frac{\max\left\{ b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i, 0 \right\}}{\|\boldsymbol{a}(\boldsymbol{x})\|_*}, \min_{i\in\mathcal{I}(\boldsymbol{x})} \chi_{\{\boldsymbol{x}: b_i(\boldsymbol{x})<0\}}(\boldsymbol{x}) \right\}, \tag{18}$$

$\mathcal{I}(\boldsymbol{x}) = [I]$ if $\boldsymbol{a}(\boldsymbol{x}) \neq 0$ and $\mathcal{I}(\boldsymbol{x}) = \emptyset$, otherwise, and characteristic function $\chi_{\mathcal{R}}(\boldsymbol{x}) = \infty$ if $\boldsymbol{x} \notin \mathcal{R}$ and 0, otherwise.

Note that the result in Theorem 7 can be further simplified by reformulating set Z as a disjunction of a nonconvex set and a convex set.

Theorem 8. *Suppose Assumptions (A1)-(A2) hold. Then $Z = Z_1 \cup Z_2$, where*

$$Z_1 = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \delta\nu - \epsilon\gamma \leq \frac{1}{N} \sum_{j \in [N]} z_j, & \text{(19a)} \\ z_j + \gamma \leq \max\left\{ b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j, 0 \right\}, \forall i \in [I], j \in [N], & \text{(19b)} \\ z_j \leq 0, \forall j \in [N], & \text{(19c)} \\ \|\boldsymbol{a}(\boldsymbol{x})\|_* \leq \nu, & \text{(19d)} \\ \nu > 0, \gamma \geq 0, & \text{(19e)} \end{array} \right\}$$

and

$$Z_2 = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}(\boldsymbol{x}) = 0, b_i(\boldsymbol{x}) \geq 0, \forall i \in [I] \right\}. \tag{20}$$

Proof. We need to show that $Z_1 \cup Z_2 \subseteq Z$ and $Z \subseteq Z_1 \cup Z_2$.

$Z_1 \cup Z_2 \subseteq Z$. Given $\boldsymbol{x} \in Z_2$, we have $\mathcal{I}(\boldsymbol{x}) = [I]$, thus $f(\boldsymbol{x}, \boldsymbol{\zeta})$ (defined in (18)) is ∞. Thus, let $\gamma = \frac{\delta}{\epsilon}$. Clearly, $(\gamma, \boldsymbol{x})$ satisfies all the constraints in (17), i.e., $\boldsymbol{x} \in Z$. Hence, $Z_2 \subseteq Z$.

Given $\boldsymbol{x} \in Z_1$, there exists $(\gamma, \nu, \boldsymbol{z}, \boldsymbol{x})$ which satisfies constraints in (19). Suppose that $\mathcal{I}(\boldsymbol{x}) = [I]$, then we have $\boldsymbol{a}(\boldsymbol{x}) = 0$. Hence, for each $i \in \mathcal{I}(\boldsymbol{x})$, we have (19a) and (19b) imply that

$$\delta\nu - \epsilon\gamma \leq \frac{1}{N} \sum_{j \in [N]} z_j \leq \frac{1}{N} \sum_{j \in [N]} (\max\{b_i(\boldsymbol{x}), 0\} - \gamma),$$

which is equivalent to

$$\max\{b_i(\boldsymbol{x}), 0\} \geq \delta\nu + (1 - \epsilon)\gamma > 0.$$

That is, $b_i(\boldsymbol{x}) > 0$. Thus, $\boldsymbol{x} \in Z_2 \subseteq Z$.

Now we suppose that $\mathcal{I}(\boldsymbol{x}) = \emptyset$. For each $i \in [I]$, (19a) and (19b) along with $\nu > 0$ imply that

$$\begin{aligned} \frac{z_j}{\nu} &\leq \min\left\{ \frac{1}{\nu} \min_{i \in [I]} \max\left\{ b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j, 0 \right\} - \frac{\gamma}{\nu}, 0 \right\} \\ &\leq \min\left\{ \min_{i \in [I]} \frac{\max\left\{ b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j, 0 \right\}}{\|\boldsymbol{a}(\boldsymbol{x})\|_*} - \frac{\gamma}{\nu}, 0 \right\} \\ &= \min\left\{ f(\boldsymbol{x}, \boldsymbol{\zeta}^j) - \frac{\gamma}{\nu}, 0 \right\} \end{aligned}$$

where the second inequality due to (19d). Then according to (19a), we have

$$\delta - \epsilon\frac{\gamma}{\nu} \leq \frac{1}{N} \sum_{j \in [N]} \frac{z_j}{\nu} \leq \frac{1}{N} \sum_{j \in [N]} \min\left\{ f(\boldsymbol{x}, \boldsymbol{\zeta}^j) - \frac{\gamma}{\nu}, 0 \right\}$$

i.e., $(\gamma/\nu, \boldsymbol{x})$ satisfies the constraints in (17), i.e., $\boldsymbol{x} \in Z$. Thus, $Z_1 \subseteq Z$.

$Z \subseteq Z_1 \cup Z_2$ Similarly, given $\boldsymbol{x} \in Z$, there exists $(\gamma, \boldsymbol{x})$ which satisfies constraints in (17). Suppose that $\boldsymbol{a}(\boldsymbol{x}) = 0$, then we must have $b_i(\boldsymbol{x}) \geq 0$ for all $i \in [I]$, otherwise, we have $f(\boldsymbol{x}, \boldsymbol{\zeta}^j) = 0$ for all $j \in [I]$. Then (17a) is equivalent to

$$0 < \delta \leq (\epsilon - 1)\gamma$$

a contradiction that $\gamma \geq 0, \epsilon \in (0,1)$. Hence, we must $\boldsymbol{x} \in Z_2$.

From now on, we assume that $\boldsymbol{a}(\boldsymbol{x}) \neq 0$. Let us define $\widehat{\gamma} = \gamma \|\boldsymbol{a}(\boldsymbol{x})\|_*, \nu = \|\boldsymbol{a}(\boldsymbol{x})\|_*$, and $z_j = \min_{i \in [I]}(\max\{b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j, 0\} - \widehat{\gamma}, 0)$ for each $j \in [N]$. Clearly, $(\widehat{\gamma}, \nu, \boldsymbol{z}, \boldsymbol{x})$ satisfies constraints in (19), i.e., $\boldsymbol{x} \in Z_1$.

□

We remark that set Z_2 is usually trivial.

Remark 1. (i) If $\eta = 1$, then

$$Z_2 = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} = 0, b_i \geq 0, \forall i \in [I]\};$$

(ii) if $\eta = 0$, then $Z_2 = \emptyset$.

On the other hand, set Z_1 can be formulated as a mixed integer set when it is bounded, i.e., we can introduce binary variables to represent constraints (19b).

Theorem 9. *Suppose there exists an $M \in \mathbb{R}_+^{I \times N}$ such that*

$$\max_{\boldsymbol{x} \in Z_1} |b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j| \leq M_{ij}$$

for all $i \in [I], j \in [N]$. Then Z_1 is mixed integer representable as below:

$$Z_1 = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \delta\nu - \epsilon\gamma \leq \frac{1}{N} \sum_{j \in [N]} z_j, & \text{(21a)} \\ z_j + \gamma \leq s_{ij}, \forall i \in [I], j \in [N], & \text{(21b)} \\ s_{ij} \geq b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j, \forall i \in [I], j \in [N], & \text{(21c)} \\ s_{ij} \leq M_{ij} y_{ij}, s_{ij} \leq b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j + M_{ij}(1 - y_{ij}), \forall i \in [I], j \in [N], & \text{(21d)} \\ \|\boldsymbol{a}(\boldsymbol{x})\|_* \leq \nu, & \text{(21e)} \\ \nu > 0, \gamma \geq 0, s_{ij} \geq 0, z_j \leq 0, y_{ij} \in \{0,1\}, \forall i \in [I], j \in [N]. & \text{(21f)} \end{array} \right\}$$

Proof. To prove that Z_1 is equivalent to the right-hand side of (21), it is sufficient to show that for each $i \in [I], j \in [N]$,

$$\max\left\{b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j, 0\right\} = s_{ij}.$$

There are three cases:

Case 1. if $b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j < 0$, then we must have $y_{ij} = 0$ (otherwise, we have $s_{ij} \leq b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j < 0$, a contradiction that $s_{ij} \geq 0$). Hence, we have $s_{ij} = 0 = \max\left\{b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j, 0\right\}$.

Case 2. if $b_i(\boldsymbol{x})-\boldsymbol{a}(\boldsymbol{x})^\top\boldsymbol{\zeta}_i^j=0$, then for any $y_{ij}\in\{0,1\}$, we have $s_{ij}=0=\max\left\{b_i(\boldsymbol{x})-\boldsymbol{a}(\boldsymbol{x})^\top\boldsymbol{\zeta}_i^j,0\right\}$.

Case 3. if $b_i(\boldsymbol{x})-\boldsymbol{a}(\boldsymbol{x})^\top\boldsymbol{\zeta}_i^j>0$, then we must have $y_{ij}=1$ (otherwise, we have $b_i(\boldsymbol{x})-\boldsymbol{a}(\boldsymbol{x})^\top\boldsymbol{\zeta}_i^j\leq s_{ij}\leq M_jy_j=0$, a contradiction that $s_{ij}\geq b_i(\boldsymbol{x})-\boldsymbol{a}(\boldsymbol{x})^\top\boldsymbol{\zeta}_i^j>0$). Thus, we has $s_{ij}=b_i(\boldsymbol{x})-\boldsymbol{a}(\boldsymbol{x})^\top\boldsymbol{\zeta}_i^j=\max\left\{b_i(\boldsymbol{x})-\boldsymbol{a}(\boldsymbol{x})^\top\boldsymbol{\zeta}_i^j,0\right\}$.

□

Note that there are various methods introduced by literature [30, 34] to obtain big-M coefficients and in the numerical study section, we will derive the big-M coefficients by inspection.

In formulation (21), there are $I\times N$ binary variables and big-M coefficients, causing it very challenging to solve. In the next section, we will show that set Z_1 can be approximated to arbitrary accuracy by a big-M free formulation, and a branch and cut algorithm can be used to solve the approximated formulation.

DRCCP with Right-hand Uncertainty: In this special case, we consider DRCCP with right-hand uncertainty, i.e., $\eta=0,\boldsymbol{a}(\boldsymbol{x})=\boldsymbol{e}$. We first note that by Theorem 7, set Z with $\boldsymbol{a}(\boldsymbol{x})=\boldsymbol{e}$ yields a more compact formulation.

Theorem 10. *If $\eta=0,\boldsymbol{a}(\boldsymbol{x})=\boldsymbol{e}$, then under Assumptions (A1)-(A2), set Z is equivalent to the following mathematical program:*

$$Z=\left\{\boldsymbol{x}\in\mathbb{R}^n:\begin{array}{ll}\delta-\epsilon\gamma\leq\dfrac{1}{N}\displaystyle\sum_{j\in[N]}z_j, & \text{(22a)}\\ (z_j+\gamma)\sqrt{n}\leq\max\left\{b_i(\boldsymbol{x})-\boldsymbol{e}^\top\boldsymbol{\zeta}_i^j,0\right\},\forall j\in[N],i\in[I], & \text{(22b)}\\ z_j\leq 0,\forall j\in[N],\gamma\geq 0. & \text{(22c)}\end{array}\right\}$$

Proof. The result directly follows from Theorem 7.

To reformulate set Z in (22) as a mixed integer program, we observe that without loss of generality, $\boldsymbol{e}^\top\boldsymbol{\zeta}_i^j$ is nonnegative for all $j\in[N],i\in[I]$. Indeed, suppose that $L:=\min_{j\in[N],i\in[I]}\boldsymbol{e}^\top\boldsymbol{\zeta}_i^j<0$, then we can redefine $\boldsymbol{e}^\top\boldsymbol{\zeta}_i^j:=\boldsymbol{e}^\top\boldsymbol{\zeta}_i^j-L$ and $b_i(\boldsymbol{x}):=b_i(\boldsymbol{x})-L$ for all $j\in[N],i\in[I]$.

Theorem 11. *Suppose that $\eta=0,\boldsymbol{a}(\boldsymbol{x})=\boldsymbol{e}$ and $\boldsymbol{e}^\top\boldsymbol{\zeta}_i^j\geq 0$ for all $i\in[I]$ and there exists an $\boldsymbol{M}\in\mathbb{R}_+^{I\times N}$ such that*

$$\max_{\boldsymbol{x}\in Z}b_i(\boldsymbol{x})-\boldsymbol{e}^\top\boldsymbol{\zeta}_i^j\leq M_{ij}$$

for all $j\in[N],i\in[I]$. Then set Z is

$$Z=\left\{\boldsymbol{x}\in\mathbb{R}^n:\begin{array}{ll}\delta-\epsilon\gamma\leq\dfrac{1}{N}\displaystyle\sum_{j\in[N]}z_j, & \text{(23a)}\\ (z_j+\gamma)\sqrt{n}\leq s_{ij},\forall j\in[N],i\in[I], & \text{(23b)}\\ s_{ij}\leq b_i(\boldsymbol{x})-\boldsymbol{e}^\top\boldsymbol{\zeta}_i^jy_{ij},\forall j\in[N],i\in[I], & \text{(23c)}\\ s_{ij}\leq M_{ij}y_{ij},\forall j\in[N],i\in[I], & \text{(23d)}\\ s_{ij}\geq b_i(\boldsymbol{x})-\boldsymbol{e}^\top\boldsymbol{\zeta}_i^j,\forall j\in[N],i\in[I], & \text{(23e)}\\ \gamma\geq 0,z_j\leq 0,y_{ij}\in\{0,1\},\forall j\in[N],i\in[I]. & \text{(23f)}\end{array}\right\}$$

Proof. Let $\widehat{Z}$ denote the set on right-hand side of (23), we would like to show that $\widehat{Z} = Z$.

$Z \subseteq \widehat{Z}$. Given $\boldsymbol{x} \in Z$, there exists $(\gamma, \boldsymbol{z}, \boldsymbol{x})$ which satisfies constraints (22). Now let $s_{ij} = \max\{b_i(\boldsymbol{x}) - \boldsymbol{e}^\top \boldsymbol{\zeta}_i^j, 0\}$ and $y_{ij} = 1$ if $b_i(\boldsymbol{x}) \geq \boldsymbol{e}^\top \boldsymbol{\zeta}_i^j$, 0, otherwise for each $j \in [N], i \in [I]$. We only need to show that $(\gamma, \boldsymbol{s}, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{x})$ satisfies the constraints in (23). Clearly, constraints (23a), (23b), (23d), (23e) and (23f) are satisfied, and for each $j \in [N], i \in [I]$ such that $y_{ij} = 1$, constraints (23c) are also satisfied since $s_{ij} = \max\left\{b_i(\boldsymbol{x}) - \boldsymbol{e}^\top \boldsymbol{\zeta}_i^j, 0\right\} = b_i(\boldsymbol{x}) - \boldsymbol{e}^\top \boldsymbol{\zeta}_i^j$. It remains to show that for each $j \in [N], i \in [I]$ such that $y_{ij} = 0$, constraints (23c) will be also satisfied, i.e., $s_{ij} = 0 \leq b_i(\boldsymbol{x})$.

Suppose that there exists a $i_0 \in [I]$ such that $b_{i_0}(x) < 0$. By assumption, we know that $\boldsymbol{e}^\top \boldsymbol{\zeta}_{i_0}^j \geq 0$ for all $j \in [N]$. Thus, we have $\max\left\{b_{i_0}(x) - \boldsymbol{e}^\top \boldsymbol{\zeta}_{i_0}^j, 0\right\} = 0$ for all $j \in [N]$. According to constraints (22b), we have $z_j + \gamma \leq 0$, i.e., $z_j \leq -\gamma$ for each $j \in [N]$. Substituting this inequality into constraint (22a), we finally have

$$\delta - \epsilon\gamma \leq \frac{1}{N} \sum_{j \in [N]} z_j \leq -\gamma$$

which implies that $\gamma \leq -\frac{\delta}{1-\epsilon} < 0$, a contradiction that $\gamma \geq 0$.

$Z \supseteq \widehat{Z}$. Given $\boldsymbol{x} \in Z$, there exists $(\gamma, \boldsymbol{s}, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{x})$ which satisfies the constraints in (23). To prove that $(\gamma, \boldsymbol{z}, \boldsymbol{x})$ satisfies constraints (22), we only need to show that

$$s_{ij} \leq \max\left\{b_i(\boldsymbol{x}) - \boldsymbol{e}^\top \boldsymbol{\zeta}_i^j, 0\right\}$$

for each $j \in [N], i \in [I]$. Indeed, if $y_{ij} = 1$, we have $s_{ij} \leq b_i(\boldsymbol{x}) - \boldsymbol{e}^\top \boldsymbol{\zeta}_i^j \leq \max\{b_i(\boldsymbol{x}) - \boldsymbol{e}^\top \boldsymbol{\zeta}_i^j, 0\}$, otherwise, we have $s_{ij} \leq 0 \leq \max\{b_i(\boldsymbol{x}) - \boldsymbol{e}^\top \boldsymbol{\zeta}_i^j, 0\}$.

□

We finally remark that formulation (23) can be stronger than (21) since we only need to compute the largest upper bound of $b_i(\boldsymbol{x}) - \boldsymbol{e}^\top \boldsymbol{\zeta}_i^j$ and define it as M_{ij} rather than the largest absolute value of $b_i(\boldsymbol{x}) - \boldsymbol{e}^\top \boldsymbol{\zeta}_i^j$.

3.2 Inner and Outer Approximations

In the previous subsection, we develop exact mixed integer reformulations of set Z under various setting. However, these reformulations might be difficult to solve, especially when the number of empirical data points becomes large (i.e., N is large), there are a large number (i.e., $I \times N$) of binary variables in the reformulations. In this subsection, we will investigate compact formulations of the inner and outer approximations of set Z proposed in Section 2, which involve fewer or even zero binary variables.

VaR Outer Approximation: We first study the reformulation of the outer approximation $Z_{\mathbf{VaR}}$.

Theorem 12. *Under Assumptions (A1)-(A2), set Z is outer approximated by*

$$Z_{\mathbf{VaR}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \mathbb{P}_{\tilde{\boldsymbol{\zeta}}} \left\{ \frac{\delta}{\epsilon} \|\boldsymbol{a}(\boldsymbol{x})\|_* + \boldsymbol{a}(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}}_i \leq b_i(\boldsymbol{x}), i \in [I] \right\} \geq 1 - \epsilon \right\}. \tag{24}$$

Proof. By Theorem 3 with

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

and $\mathcal{I}(\boldsymbol{x}) = [I]$ if $\boldsymbol{a}(\boldsymbol{x}) \neq 0$, otherwise, $\mathcal{I}(\boldsymbol{x}) = \emptyset$, we have

$$Z_{\mathbf{VaR}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \mathbb{P}_{\tilde{\boldsymbol{\zeta}}} \left\{ \begin{array}{c} \frac{\max\left\{ b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}, 0 \right\}}{\|\boldsymbol{a}(\boldsymbol{x})\|_*} \geq \frac{\delta}{\epsilon}, \forall i \in [I] \setminus \mathcal{I}(\boldsymbol{x}), \\ \chi_{b_i(\boldsymbol{x}) < 0}(x) \geq \frac{\delta}{\epsilon}, \forall i \in \mathcal{I}(\boldsymbol{x}) \end{array} \right\} \geq 1 - \epsilon \right\},$$

which is equivalent to (24). □

Note that in (24), we arrive at a regular chance constrained program with discrete random vector $\tilde{\boldsymbol{\zeta}}$, which can be reformulated as mixed integer program with big-M coefficients (cf., [1, 25]). A particular interpretation of formulation (24) is that in order to enforce the robustness, we further penalize the left-hand side of uncertain constraints by the dual norm $\|\boldsymbol{a}(\boldsymbol{x})\|_*$.

Inner Approximation I- Robust Scenario Approximation: We next consider robust scenario approximation Z_S.

Theorem 13. *Under Assumptions (A1)-(A2), set Z is inner approximated by*

$$Z_S = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \frac{\delta}{\epsilon} \|\boldsymbol{a}(\boldsymbol{x})\|_* + \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}^j \leq b_i(\boldsymbol{x}), \forall j \in [N], i \in [I]. \right\} \tag{25}$$

Proof. The proof is similar to Theorem 12, thus is omitted .□

We remark that set Z_S in (25) is very similar to scenario approach to chance constrained program [6], i.e., generate N i.i.d. samples $\{\boldsymbol{\zeta}^j\}_{j \in [N]}$ and enforce the constraints corresponding to each sample. The difference is that in set Z_S, we also add a penalty $\frac{\delta}{\epsilon}\|\boldsymbol{a}(\boldsymbol{x})\|_*$ to each of the sampled constraints.

Inner Approximation II- Inner Chance Constraint Approximation: The second inner approximation set Z_I is nonconvex and according to Theorem 5, we can formulate it as below.

Theorem 14. *Suppose that Assumptions (A1)-(A2) hold and $\epsilon \in (0, 1), N\epsilon \notin \mathbb{Z}_+$, then set Z is inner approximated by*

$$Z_I = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \mathbb{P}_{\tilde{\boldsymbol{\zeta}}} \left\{ \frac{\delta}{\epsilon - \alpha} \|\boldsymbol{a}(\boldsymbol{x})\|_* + \boldsymbol{a}(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}}_i \leq b_i(\boldsymbol{x}), i \in [I] \right\} \geq 1 - \alpha, 0 \leq \alpha \leq \frac{\lfloor N\epsilon \rfloor}{N} \right\}. \tag{26}$$

Proof. The proof is similar to Theorem 12, hence is omitted.

Note that for any given α, set Z_I is mixed integer representable with big-M coefficients. Since from Corollary 2, there are only $\lfloor N\epsilon \rfloor + 1$ effective values of α that we can choose from, thus Z_I can be formulated as a disjunction of $\lfloor N\epsilon \rfloor + 1$ mixed integer sets.

Inner Approximation III- CVaR Approximation: Next, we study the CVaR approximation.

Theorem 15. *Under Assumptions (A1)-(A2), set Z is inner approximated by*

$$Z_{\mathbf{CVaR}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \delta\nu - \epsilon\gamma \leq \dfrac{1}{N} \displaystyle\sum_{j\in[N]} z_j, & (27a) \\ z_j + \gamma \leq b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j, \forall j \in [N], i \in [I], & (27b) \\ z_j \leq 0, \forall j \in [N], & (27c) \\ \|\boldsymbol{a}(\boldsymbol{x})\|_* \leq \nu, , & (27d) \\ \nu \geq 0, \gamma \geq 0, & (27e) \end{array} \right\}$$

Proof. By Theorem 6, we have set $Z_{\mathbf{CVaR}}$ is equal to

$$Z_{\mathbf{CVaR}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} -\epsilon\beta + \lambda\delta - \dfrac{1}{N} \displaystyle\sum_{j\in[N]} \inf_{\boldsymbol{\xi}} \left[\lambda \|\boldsymbol{\xi} - \boldsymbol{\zeta}^j\| - \left(\max_{i\in[I]} \left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i - b_i(\boldsymbol{x}) \right) + \beta \right)_+ \right] \leq 0, \\ \lambda, \beta \geq 0. \end{array} \right\}$$

which is further equivalent to

$$Z_{\mathbf{CVaR}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} -\epsilon\beta + \lambda\delta - \dfrac{1}{N} \displaystyle\sum_{j\in[N]} \min \left[\min_{i\in[I]} \left(b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j \right) - \beta, 0 \right] \leq 0, \\ \|\boldsymbol{a}(\boldsymbol{x})\|_* \leq \lambda, \forall i \in [I], \\ \lambda, \beta \geq 0. \end{array} \right\}$$

In the above formulation, let $\nu = \lambda, \gamma = \beta$ and also let $z_j = \min \left[\min_{i\in[I]} \left(b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j \right) - \beta, 0 \right]$ and linearize it for each $j \in [N]$. Thus, we arrive at (27).□

We remark that we can directly derive the reformulation of set $Z_{\mathbf{CVaR}}$ in (27) based upon formulation (17). Indeed, since $\max\{b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i, 0\} \geq b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i$, by replacing $\max\{b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i, 0\}$ with $b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}$, then function $f(\boldsymbol{x}, \boldsymbol{\zeta})$ is lower bounded by

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

Thus, set Z can be inner approximated by the following set

$$\left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \delta - \epsilon\gamma \leq \dfrac{1}{N} \displaystyle\sum_{j\in[N]} \min \left\{ \underline{f}(\boldsymbol{x}, \boldsymbol{\zeta}^j) - \gamma, 0 \right\}, \\ \gamma \geq 0, \end{array} \right\}$$

which is exactly equivalent to $Z_{\mathbf{CVaR}}$ by introducing additional variables to linearize the nonlinear function $\min\left\{\underline{f}(\boldsymbol{x}, \boldsymbol{\zeta}) - \gamma, 0\right\}$. From this observation, we note that $Z_{\mathbf{CVaR}} = Z$ if $N\epsilon \leq 1$.

Corollary 3. *Suppose that Assumptions (A1)-(A2) hold and $\epsilon \in (0, 1/N]$, then set $Z = Z_{\mathbf{CVaR}}$.*

Proof. We note that set $Z_{\mathbf{CVaR}} \subseteq Z = Z_1 \cup Z_2$, where Z_1 and Z_2 are defined in (19) and (20), respectively. We note that set $Z_2 \subseteq Z_{\mathbf{CVaR}}$. Indeed, suppose that $\boldsymbol{x} \in Z_2$, i.e., $\boldsymbol{a}(\boldsymbol{x}) = 0, b_i(\boldsymbol{x}) \geq 0$ for each $i \in [I]$, then let $\nu = 0, \gamma = 0$ and $z_j = 0$ for each $j \in [N]$. Clearly, $(\nu, \gamma, \boldsymbol{z}, \boldsymbol{x})$ satisfies the constraints in (27). Hence, $\boldsymbol{x} \in Z_{\mathbf{CVaR}}$.

Thus, it is sufficient to show that $Z_1 \subseteq Z_{\mathbf{CVaR}}$. Indeed, given $\boldsymbol{x} \in Z_1$, there exists $(\nu, \gamma, \boldsymbol{z})$ such that $(\nu, \gamma, \boldsymbol{z}, \boldsymbol{x})$ satisfies the constraints in (19). We only need to show that $z_j + \gamma > 0$ for each $j \in [N]$. Suppose that there exists a $j_0 \in [N]$ such that $z_{j_0} + \gamma \leq 0$. Then according to (19a), we have

$$\delta \leq \frac{1}{N} \sum_{j \in [N] \setminus \{j_0\}} z_j + \frac{1}{N}(N\epsilon\gamma + z_{j_0}) \leq 0$$

where the second inequality is due to $\epsilon N \leq 1$ and $z_{j_0} + \gamma \leq 0$, a contradiction that $\delta > 0$. Therefore, in (19b), we must have

$$\max\left\{b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j, 0\right\} = b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j$$

for each $i \in [I], j \in [N]$. Hence, $(\nu, \gamma, \boldsymbol{z}, \boldsymbol{x})$ satisfies the constraints in (27), i.e., $\boldsymbol{x} \in Z_{\mathbf{CVaR}}$. $\square$

Formulation Comparisons: Finally, we would like to compare sets $Z_S, Z_{\mathbf{CVaR}}$. Indeed, we can show that $Z_S \subseteq Z_{\mathbf{CVaR}}$, i.e., set Z_S is at least as conservative as set $Z_{\mathbf{CVaR}}$.

Theorem 16. *Let $Z_S, Z_{\mathbf{CVaR}}$ be defined in* (25), (27) *, respectively. Then*

$$Z_S \subseteq Z_{\mathbf{CVaR}}.$$

Proof. Given $\boldsymbol{x} \in Z_S$, we only need to show that $\boldsymbol{x} \in Z_{\mathbf{CVaR}}$. Indeed, let us consider $\nu = \|\boldsymbol{a}(\boldsymbol{x})\|_*$, $\gamma = \frac{\|\boldsymbol{a}(\boldsymbol{x})\|_*}{\epsilon}, z_j = 0$ for all $j \in [N]$, then we see that $(\nu, \gamma, \boldsymbol{z}, \boldsymbol{x})$ satisfies the constraints in (27), i.e., $\boldsymbol{x} \in Z_{\mathbf{CVaR}}$.$\square$

We illustrate sets $Z, Z_{\mathbf{VaR}}, Z_{\mathbf{CVaR}}, Z_S, Z_I$ with the following example.

Example 1. Suppose $N = 3, n = 2, I = 2, \delta = 1/6, \epsilon = 2/3$ and $\boldsymbol{\zeta}_1^1 = (1, 0)^\top, \boldsymbol{\zeta}_2^1 = (0, 3)^\top, \boldsymbol{\zeta}_1^2 = (3, 0)^\top, \boldsymbol{\zeta}_2^2 = (0, 1)^\top, \boldsymbol{\zeta}_1^3 = (2, 0)^\top, \boldsymbol{\zeta}_2^3 = (0, 2)^\top, b_1(x) = x_1, \boldsymbol{a}(\boldsymbol{x}) = 1, b_2(x) = x_2$. Then, we have

$$\begin{aligned}
Z &= \left\{(x_1, x_2) : 2 + \frac{\sqrt{2}}{2} \leq x_1, 3 + \frac{\sqrt{2}}{2} \leq x_2\right\} \cup \left\{(x_1, x_2) : 3 + \frac{\sqrt{2}}{2} \leq x_1, 2 + \frac{\sqrt{2}}{2} \leq x_2\right\} \\
&\quad \cup \left\{(x_1, x_2) : 3 \leq x_1, 3 \leq x_3, 6 + \frac{\sqrt{2}}{2} \leq x_1 + x_2\right\} \\
Z_{\mathbf{VaR}} &= \left\{(x_1, x_2) : 2 + \frac{\sqrt{2}}{4} \leq x_1, 3 + \frac{\sqrt{2}}{4} \leq x_2\right\} \cup \left\{(x_1, x_2) : 3 + \frac{\sqrt{2}}{4} \leq x_1, 2 + \frac{\sqrt{2}}{4} \leq x_2\right\} \\
Z_{\mathbf{CVaR}} &= \left\{(x_1, x_2) : 3 \leq x_1, 3 \leq x_2, 6 + \frac{\sqrt{2}}{2} \leq x_1 + x_2\right\}
\end{aligned}$$

$$
\begin{aligned}
Z_S &= \left\{(x_1,x_2) : 3+\frac{\sqrt{2}}{4} \le x_1, 3+\frac{\sqrt{2}}{4} \le x_2\right\} \\
Z_I &= \left\{(x_1,x_2) : 2+\frac{\sqrt{2}}{2} \le x_1, 3+\frac{\sqrt{2}}{2} \le x_2\right\} \cup \left\{(x_1,x_2) : 3+\frac{\sqrt{2}}{2} \le x_1, 2+\frac{\sqrt{2}}{2} \le x_2\right\} \\
&\quad \cup \left\{(x_1,x_2) : 3+\frac{\sqrt{2}}{4} \le x_1, 3+\frac{\sqrt{2}}{4} \le x_2\right\}.
\end{aligned}
$$

Clearly, we have $Z_S \subsetneq \left\{\begin{matrix} Z_{\mathbf{CVaR}} \\ \not\subseteq\!\!\not\supseteq \\ Z_I \end{matrix}\right\} \subsetneq Z \subsetneq Z_{\mathbf{VaR}}$ (see Figure 1 for an illustration).

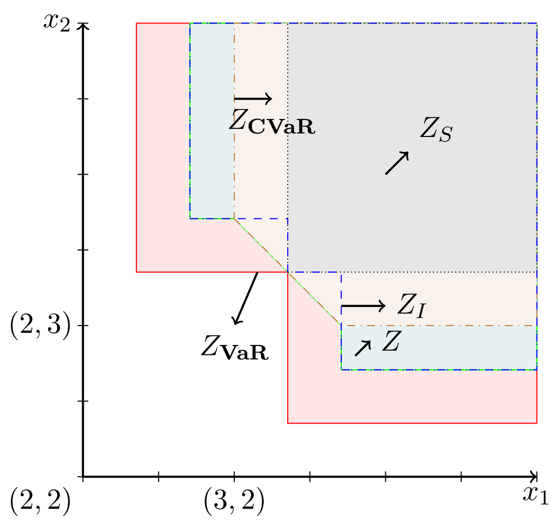

Figure 1: Illustration of Example 1

Finally, the inclusive relationships among sets $Z, Z_{\mathbf{VaR}}, Z_S, Z_I Z_{\mathbf{CVaR}}$ are illustrated in Figure 2 and their reformulations are summarized in Table 1.

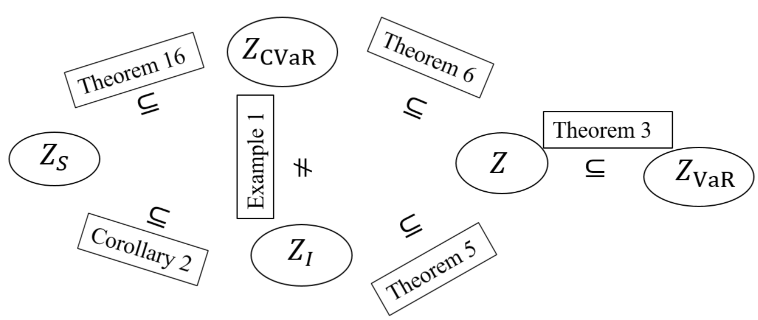

Figure 2: Summary of formulation comparisons

Table 1: Summary of formulation results in Section 3

Set Z	Set $Z_{\mathbf{VaR}}$	Set Z_S	Set Z_I	Set $Z_{\mathbf{CVaR}}$
Mixed-integer	Mixed-integer	Convex	Mixed-integer	Convex
Theorem 9	Theorem 12	Theorem 13	Theorem 14	Theorem 15

4 DRCCP with Pure Binary Decision Variables

In this section, we will study DRCCP with pure binary decision variables $\boldsymbol{x} \in \{0,1\}^n$, i.e., in addition to Assumptions (A1)-(A2), we further assume that

(A3) the set S is binary, that is, $S \subseteq \{0,1\}^n$.

Indeed, we remark that if S is a bounded mixed integer set, we can introduce $O(n\log(n))$ additional binary variables to approximate set S with arbitrary accuracy via binary approximation of continuous variables (c.f., [42]). For binary DRCCP, we will show that the reformulations in the previous section can be improved.

4.1 Polyhedral Results of Submodular Functions: A Review

Our main derivation of stronger formulations is based upon some polyhedral results of submodular functions, which will be briefly reviewed in this subsection.

We first begin with the following lemmas on submodular functions.

Lemma 1. *Given $\boldsymbol{d}_1 \in \mathbb{R}^n_+, d_2, d_3 \in \mathbb{R}$, function $f(\boldsymbol{x}) = -\max\left(\boldsymbol{d}_1^\top \boldsymbol{x} + d_2, d_3\right)$ is submodular over the binary hypercube.*

Proof. For simplicity, given a $T \subseteq [n]$, we define a binary vector $\boldsymbol{e}_T \in \{0,1\}^n$ such that

$$(\boldsymbol{e}_T)_l = \begin{cases} 1, & \text{if } l \in T \\ 0, & \text{if } l \in [n] \setminus T \end{cases}.$$

According to the definition of submodular function [12], we only need to show that

$$f(\boldsymbol{e}_{T_1 \cup \{t\}}) - f(\boldsymbol{e}_{T_1}) \geq f(\boldsymbol{e}_{T_2 \cup \{t\}}) - f(\boldsymbol{e}_{T_2})$$

for any $T_1 \subseteq T_2$ and $t \in [n] \setminus T_2$. There are three cases:

Case 1. if $\sum_{i \in T_1} d_{1i} + d_2 \geq d_3$, since $\boldsymbol{d}_1 \in \mathbb{R}^n_+$, then we must have

$$f(\boldsymbol{e}_{T_1 \cup \{t\}}) - f(\boldsymbol{e}_{T_1}) = f(\boldsymbol{e}_{T_2 \cup \{t\}}) - f(\boldsymbol{e}_{T_2}) = -d_{1t}.$$

Case 2. if $\sum_{i \in T_1} d_{1i} + d_2 < d_3$ but $\sum_{i \in T_2} d_{1i} + d_2 \geq d_3$, then we must have

$$f(\boldsymbol{e}_{T_1 \cup \{t\}}) - f(\boldsymbol{e}_{T_1}) = 0 \geq f(\boldsymbol{e}_{T_2 \cup \{t\}}) - f(\boldsymbol{e}_{T_2}) = -d_{1t},$$

where the inequality is due to $\boldsymbol{d}_1 \in \mathbb{R}^n_+$.

Case 3. if $\sum_{i \in T_2} d_{1i} + d_2 < d_3$, since $\boldsymbol{d}_1 \in \mathbb{R}^n_+$, then we must have

$$f(\boldsymbol{e}_{T_1 \cup \{t\}}) - f(\boldsymbol{e}_{T_1}) = f(\boldsymbol{e}_{T_2 \cup \{t\}}) - f(\boldsymbol{e}_{T_2}) = 0.$$

□

Lemma 2. *Given $q \geq 1$, function $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_q$ with $q \geq 1$ is submodular over the binary hypercube.*

Proof. This is because $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_q = \sqrt[q]{\sum_{l \in [n]} x_l}$, and $g(\boldsymbol{e}^\top \boldsymbol{x})$ is a submodular function if $g(\cdot)$ is a concave function (cf., [40]). □

Next, we will introduce polyhedral properties of submodular functions. For any given submodular function $f(\boldsymbol{x})$ with $\boldsymbol{x} \in \{0,1\}^n$, let us denote Π_f to be its epigraph, i.e.,

$$\Pi_f = \{(\boldsymbol{x}, \phi) : \phi \geq f(\boldsymbol{x}), \boldsymbol{x} \in \{0,1\}^n\}.$$

Then the convex hull of Π_f is characterized by the system of *"extended polymatroid inequalities"* (EPI) [2, 40], i.e.,

$$\operatorname{conv}(\Pi_f) = \left\{ (\boldsymbol{x}, \phi) : f(\boldsymbol{0}) + \sum_{l \in [n]} \rho_{\sigma_l} x_{\sigma_l} \leq \phi, \forall \sigma \in \Omega, \boldsymbol{x} \in [0,1]^n \right\}, \tag{28}$$

where Ω denotes a collection of all the permutations of set $[n]$ and $\rho_{\sigma_l} = f(\boldsymbol{e}_{A_l^\sigma}) - f(\boldsymbol{e}_{A_{l-1}^\sigma})$ for each $l \in [n]$ with $A_0^\sigma = \emptyset, A_l^\sigma = \{\sigma_1, \ldots, \sigma_l\}$ and $(\boldsymbol{e}_T)_\tau = \begin{cases} 1, & \text{if } \tau \in T \\ 0, & \text{if } \tau \in [n] \setminus T \end{cases}$.

In addition, although there are $n!$ number of inequalities in (28), these inequalities can be easily separated by greedy procedure.

Lemma 3. *([2, 40]) Suppose $(\tilde{\boldsymbol{x}}, \tilde{\phi}) \notin \operatorname{conv}(\Pi_f)$, and $\sigma \in \Omega$ be a permutation of $[n]$ such that $\tilde{x}_{\sigma_1} \geq \ldots \geq \tilde{x}_{\sigma_n}$. Then $(\tilde{\boldsymbol{x}}, \tilde{\phi})$ must violate the constraint $f(\boldsymbol{0}) + \sum_{l \in [n]} \rho_{\sigma_l} x_{\sigma_l} \leq \phi$.*

From Lemma 3, we see that to separate a point $(\tilde{\boldsymbol{x}}, \tilde{\phi})$ from $\operatorname{conv}(\Pi_f)$, we only need to sort the coordinates of $\tilde{\boldsymbol{x}}$ in a descending order, i.e., $\tilde{x}_{\sigma_1} \geq \ldots \geq \tilde{x}_{\sigma_n}$. Then $(\tilde{\boldsymbol{x}}, \tilde{\phi})$ can be the separated by the constraint $f(\boldsymbol{0}) + \sum_{l \in [n]} \rho_{\sigma_l} x_{\sigma_l} \leq \phi$ from $\operatorname{conv}(\Pi_f)$. The time complexity of this separating procedure is $O(n \log n)$.

4.2 Reformulating a Binary DRCCP by Submodular Knapsack Constraints: Big-M free

In this section, we will replace the nonlinar constraints defining the feasible region of a binary DRCCP (i.e., set $S \cap Z$) by submodular upper bound (knapsack) constraints. These constraints can be equivalently described by the system of EPI in (28), therefore we obtain a big-M free mixed integer representation of set $S \cap Z$.

First, we introduce n complementing binary variables of $\boldsymbol{x}$, denoted by $\boldsymbol{w}$, i.e., $w_l + x_l = 1$ for each $l \in [n]$. With these n additional variables, we can reformulate function $b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j$ as

$$b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j = \boldsymbol{r}_{ij}^\top \boldsymbol{x} + \boldsymbol{t}_{ij}^\top \boldsymbol{w} + u_{ij} \tag{29}$$

for each $i \in [I], j \in [N]$ such that $\boldsymbol{r}_{ij} \in \mathbb{R}^n_+, \boldsymbol{t}_{ij} \in \mathbb{R}^n_+$. Indeed, since $\boldsymbol{a}(\boldsymbol{x}) = \eta \boldsymbol{x} + (1-\eta)\boldsymbol{e}$ and $b_i(\boldsymbol{x}) = B_i^\top \boldsymbol{x} + b^i$, in (29), we can choose

$$
\begin{aligned}
r_{ijl} &= B_{il}\mathbb{I}(B_{il} > 0) - \eta \zeta^j_{il}\mathbb{I}(\zeta^j_{il} < 0),\\
t_{ijl} &= -B_{il}\mathbb{I}(B_{il} < 0) + \eta \zeta^j_{il}\mathbb{I}(\zeta^j_{il} > 0),\\
u_{ij} &= b^i - (1-\eta)\boldsymbol{e}^\top \boldsymbol{\zeta}^j_i + \sum_{\tau \in [n]} \left(B_{i\tau}\mathbb{I}(B_{i\tau} < 0) - \eta \zeta^j_{i\tau}\mathbb{I}(\zeta^j_{i\tau} > 0) \right),
\end{aligned}
$$

for each $l \in [n], i \in [I], j \in [N]$.

Thus, from above discussion, we can formulate $S \cap Z_1$ (note that set $Z = Z_1 \cup Z_2$ according to Theorem 8) as the following mixed integer set with submodular knapsack constraints.

Theorem 17. *Suppose that Assumptions (A1)-(A3) hold. Then $S \cap Z = (S \cap Z_1) \cup (S \cap Z_2)$, where*

$$
S \cap Z_1 = \left\{ \boldsymbol{x} \in S : \begin{array}{ll}
\delta\nu - \epsilon\gamma \le \frac{1}{N}\sum_{j \in [N]} z_j, & \text{(30a)}\\
-\max\left\{ \boldsymbol{r}_{ij}^\top \boldsymbol{x} + \boldsymbol{t}_{ij}^\top \boldsymbol{w} + u_{ij}, 0 \right\} \le -z_j - \gamma, \forall i \in [I], j \in [N], & \text{(30b)}\\
z_j \le 0, \forall j \in [N], & \text{(30c)}\\
\eta\|\boldsymbol{x}\|_* + (1-\eta)\|\boldsymbol{e}\|_* \le \nu, & \text{(30d)}\\
w_l + x_l = 1, \forall l \in [n], & \text{(30e)}\\
\nu \ge 1, & \text{(30f)}\\
\gamma \ge 0, \boldsymbol{w} \in \{0,1\}^n & \text{(30g)}
\end{array} \right\}
$$

and

$$
S \cap Z_2 = \{\boldsymbol{x} \in S : \boldsymbol{a}(\boldsymbol{x}) = 0, b_i(\boldsymbol{x}) \ge 0, \forall i \in [I]\} \tag{31}
$$

Proof. From the discussion above and the fact that $\boldsymbol{a}(\boldsymbol{x}) = \eta \boldsymbol{x} + (1-\eta)\boldsymbol{e}$ with $\eta \in \{0,1\}$, we have constraints (19b) and (19d) is equivalent to (30b) and (30d). We only need to show that $\nu \ge 1$ in set $S \cap Z_1$.

If $\eta = 0$, then $\eta\|\boldsymbol{x}\|_* + (1-\eta)\|\boldsymbol{e}\|_* = \|\boldsymbol{e}\|_* \ge 1$, then we are done.

Now suppose that $\eta = 1$. We note that if $\boldsymbol{x} = 0$, then the constraints (19) imply that $b_i(\boldsymbol{x}) > 0$ for each $i \in [I]$. Thus, if $\boldsymbol{x} = 0$, then set $Z_1 \subseteq Z_2$. Therefore, without loss of generality, we can assume that in set Z_1, $\boldsymbol{x} \neq 0$. Note that $S \cap Z_1 \subseteq \{0,1\}^n$, therefore, $\boldsymbol{x} \neq 0$ implies that $\|\boldsymbol{x}\|_* \ge 1$, thus, $v \ge \eta\|\boldsymbol{x}\|_* + (1-\eta)\|\boldsymbol{e}\|_* = \|\boldsymbol{x}\|_* \ge 1$.□

From the proof of Theorem 17, we note that if $\eta = 1$ and $b^i \ge \frac{\delta}{\epsilon}$ for each $i \in [I]$, then we have $Z_2 \subseteq Z_1$

Corollary 4. *Suppose that Assumptions (A1)-(A3) hold, $\eta = 1$ and $b^i \ge \frac{\delta}{\epsilon}$ for each $i \in [I]$. Then $S \cap Z = S \cap Z_1$.*

Proof. We only need to show that $0 \in S \cap Z_1$. Suppose $\boldsymbol{x} = 0$, i.e., $\boldsymbol{w} = \boldsymbol{e}$. Let us set $\nu = 1, \gamma = 0, \boldsymbol{z} = 0$. Then it is easy to see that $(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{z}, \gamma, \nu)$ satisfies the constraints in (30), i.e., $0 \in S \cap Z_1$. □

We note that the left-hand sides of constraints (30b) and (30d) are submodular functions according to Lemma 1 and Lemma 2, thus, we can equivalently replace these constraints with the convex hulls of epigraphs of their associated submodular functions. Thus,

Corollary 5. *Suppose that Assumptions (A1)-(A3) hold. Then*

$$S \cap Z_1 = \left\{ \boldsymbol{x} \in S : \begin{array}{ll} \delta\nu - \epsilon\gamma \leq \frac{1}{N}\sum_{j\in[N]} z_j, & \text{(32a)} \\ (\boldsymbol{x}, \boldsymbol{w}, -z_j - \gamma) \in \operatorname{conv}(\Pi_{ij}), \forall i \in [I], j \in [N], & \text{(32b)} \\ z_j \leq 0, \forall j \in [N], & \text{(32c)} \\ (\boldsymbol{x}, \nu) \in \operatorname{conv}(\Pi_0), & \text{(32d)} \\ w_l + x_l = 1, \forall l \in [n], & \text{(32e)} \\ \nu \geq 1, \gamma \geq 0, \boldsymbol{w} \in \{0,1\}^n, & \text{(32f)} \end{array} \right\}$$

where

$$\Pi_{ij} = \left\{ (\boldsymbol{x}, \boldsymbol{w}, \phi) : -\max\left\{ \boldsymbol{r}_{ij}^\top \boldsymbol{x} + \boldsymbol{t}_{ij}^\top \boldsymbol{w} + u_{ij}, 0 \right\} \leq \phi, \boldsymbol{x}, \boldsymbol{w} \in \{0,1\}^n \right\}, \forall i \in [I], j \in [N], \tag{33a}$$

$$\Pi_0 = \{ (\boldsymbol{x}, \phi) : \eta \|\boldsymbol{x}\|_* + (1-\eta)\|\boldsymbol{e}\|_* \leq \phi, \boldsymbol{x} \in \{0,1\}^n \} \tag{33b}$$

and $\{\operatorname{conv}(\Pi_{ij})\}_{i\in[I], j\in[N]}, \operatorname{conv}(\Pi_0)$ *can be described by the system of EPI in* (28).

Note that the optimization problem $\min_{\boldsymbol{x}\in S\cap Z_1} \boldsymbol{c}^\top \boldsymbol{x}$ can be solved by branch and cut algorithm. In particular, at each branch and bound node, denoted as $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{w}}, \widehat{\boldsymbol{z}}, \widehat{\gamma}, \widehat{\nu})$, there might be too many (i.e., $N \times I + 1$) valid inequalities to add, since in (32b) and (32d), there are $N \times I + 1$ convex hulls of epigraphs (i.e., $\{\operatorname{conv}(\Pi_{ij})\}_{i\in[I], j\in[N]}, \operatorname{conv}(\Pi_0)$) to be separated from. Therefore, instead, we can first check and find the epigraphs of κ ($\kappa \geq 1$) most violated constraints in (30b) and (30d), i.e., find the epigraphs corresponding to the κ largest values in the following set

$$\{-\max\left\{ \boldsymbol{r}_{ij}^\top \widehat{\boldsymbol{x}} + \boldsymbol{t}_{ij}^\top \widehat{\boldsymbol{w}} + u_{ij}, 0 \right\} + \widehat{z}_j + \widehat{\gamma}\}_{i\in[I], j\in[N]} \cup \{\eta \|\widehat{\boldsymbol{x}}\|_* + (1-\eta)\|\boldsymbol{e}\|_* - \widehat{\nu}\}.$$

Finally, we can generate and add valid inequalities by separating $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{w}}, \widehat{\boldsymbol{z}}, \widehat{\gamma}, \widehat{\nu})$ from the convex hulls of these κ epigraphs according to Lemma 3.

4.3 Numerical Study

In this subsection, we present a numerical study to compare the big-M formulation in Theorem 9 with big-M free formulation in Theorem 17 and its corollaries on the distributionally robust multidimensional knapsack problem (DRMKP) [10, 34, 37]. In DRMKP, there are n items and I knapsacks. Additionally, c_j represents the value of item j for all $j \in [n]$, $\tilde{\boldsymbol{\xi}}_i := [\tilde{\xi}_{i1}, \ldots, \tilde{\xi}_{in}]^\top$ represents the vector of random item weights in knapsack i, and $b^i > 0$ represents the capacity limit of knapsack i, for all $i \in [I]$. The binary decision variable $x_j = 1$ if the jth item is picked and 0 otherwise. We use the Wasserstein ambiguity set under Assumptions (A1) and (A2) with L_2- norm as distance metric. With the notation above, DRMKP is formulated as

$$v^* = \max_{\boldsymbol{x}\in\{0,1\}^n} \quad \boldsymbol{c}^\top \boldsymbol{x},$$

$$\text{s.t.} \quad \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left\{\tilde{\boldsymbol{\xi}}_i^\top \boldsymbol{x} \le b^i, \forall i \in [I]\right\} \ge 1-\epsilon. \tag{34}$$

To test the proposed formulations, we generate 10 random instances with $n = 20$ and $I = 10$, indexed by $\{1, 2, \ldots, 10\}$. For each instance, we generate $N = 1000$ empirical samples $\{\boldsymbol{\zeta}^j\}_{j\in[N]} \in \mathbb{R}_+^{I\times n}$ from a uniform distribution over a box $[1, 10]^{I\times n}$. For each $l \in [n]$, we independently generate c_l from the uniform distribution on the interval $[1, 10]$, while for each $i \in [I]$, we set $b^i := 100$. We test these 10 random instances with risk parameter $\epsilon \in \{0.05, 0.10\}$ and Wasserstein radius $\delta \in \{0.1, 0.2\}$.

Our first approach is to solve the big-M reformulation of DRMKP in Theorem 9, which reads as follows:

$$\begin{aligned}
v^* = \max_{\boldsymbol{x}\in\{0,1\}^n} \quad & \boldsymbol{c}^\top \boldsymbol{x}, \\
\text{s.t.} \quad & \delta\nu - \epsilon\gamma \le \frac{1}{N}\sum_{j\in[N]} z_j, \\
& z_j + \gamma \le s_{ij}, \forall i \in [I], j \in [N], \\
& s_{ij} \ge b^i - \boldsymbol{x}^\top \boldsymbol{\zeta}_i^j, \forall i \in [I], j \in [N], \\
& s_{ij} \le M_{ij} y_{ij}, s_{ij} \le b^i - \boldsymbol{x}^\top \boldsymbol{\zeta}_i^j + M_{ij}(1 - y_{ij}), \forall i \in [I], j \in [N], \\
& \|\boldsymbol{x}\|_2 \le \nu, \\
& \nu \ge 1, \gamma \ge 0, s_{ij} \ge 0, z_j \le 0, y_{ij} \in \{0, 1\}, \forall i \in [I], j \in [N],
\end{aligned} \tag{35}$$

where $M_{ij} = \max\{b^i, |b^i - \boldsymbol{e}^\top \boldsymbol{\zeta}_i^j|\}$ for each $i \in [I], j \in [N]$.

We compare this formulation with another big-M free formulation in Theorem 17 and its corollaries, which reads as follows:

$$\begin{aligned}
v^* = \max_{\boldsymbol{x}\in\{0,1\}^n} \quad & \boldsymbol{c}^\top \boldsymbol{x}, \\
\text{s.t.} \ & \delta\nu - \epsilon\gamma \le \frac{1}{N}\sum_{j\in[N]} z_j, \\
& (\boldsymbol{w}, -z_j - \gamma) \in \operatorname{conv}(\Pi_{ij}), \forall i \in [I], j \in [N], \\
& z_j \le 0, \forall j \in [N], \\
& (\boldsymbol{x}, \nu) \in \operatorname{conv}(\Pi_0), \\
& w_l + x_l = 1, \forall l \in [n], \\
& \nu \ge 1, \gamma \ge 0, \boldsymbol{w} \in [0, 1]^n,
\end{aligned} \tag{36}$$

where

$$\Pi_{ij} = \left\{(\boldsymbol{w}, \phi) : -\max\left\{(\boldsymbol{\zeta}_i^j)^\top \boldsymbol{w} + b^i - (\boldsymbol{\zeta}_i^j)^\top \boldsymbol{e}, 0\right\} \le \phi, \boldsymbol{w} \in \{0, 1\}^n\right\}, \forall i \in [I], j \in [N], \tag{37a}$$

$$\Pi_0 = \{(\boldsymbol{x}, \phi) : \|\boldsymbol{x}\|_2 \le \phi, \boldsymbol{x} \in \{0, 1\}^n\} \tag{37b}$$

and their convex hulls, $\{\operatorname{conv}(\Pi_{ij})\}_{i\in[I],j\in[N]}$, $\operatorname{conv}(\Pi_0)$ can be described by the system of EPI (28). Note that the fact that (35) and (36) are exact reformulations of DRMKP follows from Corollary 4 since $b^i \ge \frac{\delta}{\epsilon}$ for all $i \in [I]$.

We use the commercial solver Gurobi (version 7.5, with default settings) to solve the instances of formulation (35). We set the time limit of solving each instance to be 3600 seconds. The results are displayed in Table 2. We use UB, LB, GAP, Opt. Val. and Time to denote the best upper bound, the best lower bound, optimality gap, the optimal objective value and the total running time, respectively. All instances were executed on a MacBook Pro with a 2.80 GHz processor and 16GB RAM.

From Table 2, we observe that the overall running time of DRMKP formulation (36) significantly outperforms that of (35), i.e., almost all of the instances of formulation (36) can be solved within 10 minutes, while the majority of the instances of formulation (35) reach the time limit. The main reasons are two-fold: (i) formulation (35) involves $\mathcal{O}(N\times I+n)$ binary variables and $\mathcal{O}(N\times I)$ continuous variables, while formulation (35) only involves $\mathcal{O}(n)$ binary variables and $\mathcal{O}(N)$ continuous variables; and (ii) formulation (35) contains big-M coefficients, while formulation (36) is big-M free. We also observe that, as the risk parameter ϵ increases or Wasserstein radius δ decreases, both formulations take longer to solve but formulation (36) still significantly outperforms formulation (35). These results demonstrate the effectiveness our proposed approaches.

5 Conclusion

In this paper, we studied a distributionally robust chance constrained problem (DRCCP) with Wasserstein ambiguity set. We showed that a DRCCP can be formulated as a conditional-value-at-risk constrained optimization, thus admits tight inner and outer approximations. When the metric space of random variables is normed vector space, we showed that a DRCCP is mixed integer representable with big-M coefficients and additional binary variables, i.e., a DRCCP can be formulated as a mixed integer conic program. We also compared various inner and outer approximations and proved their corresponding inclusive relations. We further proposed a big-M free formulation for a binary DRCCP. The numerical studies demonstrated that the developed big-M free formulation can significantly outperform the big-M one.

Acknowledgments

The author would like to thank Professor Shabbir Ahmed (Georgia Tech) for his helpful comments on an earlier version of the paper.

Table 2: Performance comparison of formulation (35) and formulation (36)

ϵ	δ	Instances	n	I	Formulation (35)				Formulation (36)	
					UB	LB	Time	GAP	Opt. Val.	Time
0.05	0.1	1	20	10	93	86	3600.0	7.5%	89	49.3
		2	20	10	97	90	3600.0	7.2%	95	30.6
		3	20	10	95	84	3600.0	11.6%	90	387.0
		4	20	10	84	74	3600.0	11.9%	78	275.7
		5	20	10	87	81	3600.0	6.9%	82	140.4
		6	20	10	97	85	3600.0	12.4%	88	972.5
		7	20	10	89	75	3600.0	15.7%	84	169.6
		8	20	10	100	88	3600.0	12.0%	96	80.5
		9	20	10	96	78	3600.0	18.8%	92	59.3
		10	20	10	93	93	3542.7	0.0%	93	18.2
		Average					3594.3	10.4%		218.3
0.1	0.1	1	20	10	100	NA	3600.0	NA	92	172.9
		2	20	10	106	NA	3600.0	NA	99	164.0
		3	20	10	105	87	3600.0	17.1%	93	569.1
		4	20	10	92	67	3600.0	27.2%	82	600.5
		5	20	10	95	NA	3600.0	NA	86	332.0
		6	20	10	109	NA	3600.0	NA	94	1852.4
		7	20	10	96	NA	3600.0	NA	88	279.8
		8	20	10	108	82	3600.0	24.1%	100	133.2
		9	20	10	102	NA	3600.0	NA	94	389.3
		10	20	10	103	96	3600.0	6.8%	96	149.7
		Average					3600.0	18.8%		464.3
0.05	0.2	1	20	10	87	87	665.8	0.0%	87	8.5
		2	20	10	88	88	2473.2	0.0%	88	19.3
		3	20	10	86	86	1391.3	0.0%	86	70.4
		4	20	10	74	74	2881.7	0.0%	74	102.5
		5	20	10	78	78	1553.5	0.0%	78	26.9
		6	20	10	86	86	2776.2	0.0%	86	442.7
		7	20	10	83	83	1413.9	0.0%	83	17.1
		8	20	10	92	92	297.7	0.0%	92	21.0
		9	20	10	90	90	148.5	0.0%	90	14.6
		10	20	10	90	90	1074.2	0.0%	90	8.9
		Average					1467.6	0.0%		73.2
0.1	0.2	1	20	10	96	85	3600.0	11.5%	92	34.3
		2	20	10	103	88	3600.0	14.6%	99	16.5
		3	20	10	98	93	3600.0	5.1%	93	175.4
		4	20	10	86	82	3600.0	4.7%	82	243.5
		5	20	10	90	NA	3600.0	NA	86	84.7
		6	20	10	101	81	3600.0	19.8%	94	524.6
		7	20	10	90	88	3600.0	2.2%	88	93.1
		8	20	10	103	NA	3600.0	NA	100	53.4
		9	20	10	97	94	3600.0	3.1%	94	75.5
		10	20	10	99	89	3600.0	10.1%	96	14.1
		Average					3600.0	8.9%		131.5

* The NA represents that no feasible solution has been found within the time limit

References

[1] Shabbir Ahmed, James Luedtke, Yongjia Song, and Weijun Xie. Nonanticipative duality, relaxations, and formulations for chance-constrained stochastic programs. *Mathematical Programming*, 162(1-2):51–81, 2017.

[2] Alper Atamtürk and Vishnu Narayanan. Polymatroids and mean-risk minimization in discrete optimization. *Operations Research Letters*, 36(5):618–622, 2008.

[3] Jose Blanchet, Lin Chen, and Xun Yu Zhou. Distributionally robust mean-variance portfolio selection with wasserstein distances. *arXiv preprint arXiv:1802.04885*, 2018.

[4] Jose Blanchet, Yang Kang, and Karthyek Murthy. Robust wasserstein profile inference and applications to machine learning. *arXiv preprint arXiv:1610.05627*, 2016.

[5] Emmanuel Boissard et al. Simple bounds for the convergence of empirical and occupation measures in 1-wasserstein distance. *Electronic Journal of Probability*, 16:2296–2333, 2011.

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

[7] Giuseppe Carlo Calafiore and Laurent El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[8] Marco C Campi, Simone Garatti, and Maria Prandini. The scenario approach for systems and control design. *Annual Reviews in Control*, 33(2):149–157, 2009.

[9] Wenqing Chen, Melvyn Sim, Jie Sun, and Chung-Piaw Teo. From CVaR to uncertainty set: Implications in joint chance-constrained optimization. *Operations research*, 58(2):470–485, 2010.

[10] Jianqiang Cheng, Erick Delage, and Abdel Lisser. Distributionally robust stochastic knapsack problem. *SIAM Journal on Optimization*, 24(3):1485–1506, 2014.

[11] Chao Duan, Wanliang Fang, L Jiang, Li Yao, and Jun Liu. Distributionally robust chance-constrained approximate ac-opf with wasserstein metric. *IEEE Transactions on Power Systems*, 2018.

[12] Jack Edmonds. Submodular functions, matroids, and certain polyhedra. *Edited by G. Goos, J. Hartmanis, and J. van Leeuwen*, 11, 1970.

[13] L. El 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.

[14] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.

[15] Rui Gao, Xi Chen, and Anton J Kleywegt. Wasserstein distributional robustness and regularization in statistical learning. *arXiv preprint arXiv:1712.06050*, 2017.

[16] Rui Gao and Anton J Kleywegt. Distributionally robust stochastic optimization with wasserstein distance. *arXiv preprint arXiv:1604.02199*, 2016.

[17] Grani A Hanasusanto and Daniel Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *arXiv preprint arXiv:1609.07505*, 2016.

[18] Grani A Hanasusanto, Vladimir Roitch, Daniel Kuhn, and Wolfram Wiesemann. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151:35–62, 2015.

[19] Grani A Hanasusanto, Vladimir Roitch, Daniel Kuhn, and Wolfram Wiesemann. Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, 65(3):751–767, 2017.

[20] Ashish R Hota, Ashish Cherukuri, and John Lygeros. Data-driven chance constrained optimization under wasserstein ambiguity sets. *arXiv preprint arXiv:1805.06729*, 2018.

[21] Ruiwei Jiang and Yongpei Guan. Data-driven chance constrained stochastic program. *Mathematical Programming*, 158:291–327, 2016.

[22] Rüdiger Kiesel, Robin Rühlicke, Gerhard Stahl, and Jinsong Zheng. The wasserstein metric and robustness in risk management. *Risks*, 4(3):32, 2016.

[23] Jaeho Lee and Maxim Raginsky. Minimax statistical learning and domain adaptation with wasserstein distances. *arXiv preprint arXiv:1705.07815*, 2017.

[24] Bowen Li, Ruiwei Jiang, and Johanna L. Mathieu. Ambiguous risk constraints with moment and unimodality information. *Mathematical Programming*, Nov 2017.

[25] James Luedtke and Shabbir Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008.

[26] Fengqiao Luo and Sanjay Mehrotra. Decomposition algorithm for distributionally robust optimization using wasserstein metric. *arXiv preprint arXiv:1704.03920*, 2017.

[27] Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, Jul 2017.

[28] Arkadi Nemirovski and Alexander Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

[29] Arkadi Nemirovski and Alexander Shapiro. Scenario approximations of chance constraints. In *Probabilistic and randomized methods for design under uncertainty*, pages 3–47. Springer, 2006.

[30] Feng Qiu, Shabbir Ahmed, Santanu S Dey, and Laurence A Wolsey. Covering linear programming with violations. *INFORMS Journal on Computing*, 26(3):531–546, 2014.

[31] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[32] Soroosh Shafieezadeh-Abadeh, Peyman Mohajerin Esfahani, and Daniel Kuhn. Distributionally robust logistic regression. In *Advances in Neural Information Processing Systems*, pages 1576–1584, 2015.

[33] Alexander Shapiro and Anton Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17(3):523–542, 2002.

[34] Yongjia Song, James R Luedtke, and Simge Küçükyavuz. Chance-constrained binary packing problems. *INFORMS Journal on Computing*, 26(4):735–747, 2014.

[35] Weijun Xie and Shabbir Ahmed. Bicriteria approximation of chance constrained covering problems. Available at Optimization Online, 2018.

[36] Weijun Xie and Shabbir Ahmed. Distributionally robust chance constrained optimal power flow with renewables: A conic reformulation. *IEEE Transactions on Power Systems*, 33(2):1860–1867, 2018.

[37] Weijun Xie and Shabbir Ahmed. On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM Journal on Optimization*, 28(2):1151–1182, 2018.

[38] Weijun Xie, Shabbir Ahmed, and Ruiwei Jiang. Optimized bonferroni approximations of distributionally robust joint chance constraints. *Available at Optimization Online*, 2017.

[39] Wenzhuo Yang and Huan Xu. Distributionally robust chance constraints for non-linear uncertainties. *Mathematical Programming*, 155:231–265, 2016.

[40] Jiajin Yu and Shabbir Ahmed. Polyhedral results for a class of cardinality constrained submodular minimization problems. *Discrete Optimization*, 24:87–102, 2017.

[41] Chaoyue Zhao and Yongpei Guan. Data-driven risk-averse two-stage stochastic program with ζ-structure probability metrics. Available at `http://www.optimization-online.org/DB_FILE/2015/07/5014.pdf`, 2015.

[42] Jikai Zou, Shabbir Ahmed, and Xu Andy Sun. Stochastic dual dynamic integer programming. *Mathematical Programming*, pages 1–42, 2017.

[43] Steve Zymler, Daniel Kuhn, and Berç Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137:167–198, 2013.