INFORMS Optimization Society
2020 Young Researchers Prize

Expository Article: On Distributionally Robust Chance Constrained Programs with Wasserstein Distance

Weijun Xie
Department of Industrial and Systems Engineering
Virginia Tech
wxie@vt.edu

Published online by the INFORMS Optimization Society on February 22, 2021.
http://connect.informs.org/optimizationsociety/home

I am truly honored and grateful to be awarded the 2020 INFORMS Optimization Society Young Researcher Prize for the work "On Distributionally Robust Chance Constrained Program with Wasserstein Distance." I would like to thank the committee members (Prof. Sam Burer, Prof. Hande Benson, Prof. Santanu Dey, Prof. Siqian Shen) for their great effort and honoring this work. This article surveys the results of the honored paper [26].

1 Introduction

1.1 Setting

We study distributional robust chance constrained programs (DRCCPs) 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 constraints specified by the random vectors $\tilde{\boldsymbol{\xi}}_i$ supported on set $\Xi_i \subseteq \mathbb{R}^{n+1}$ 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 $i \in [I]$, $\boldsymbol{a}(\boldsymbol{x}) \in \mathbb{R}^{n+1}$ and $b_i(\boldsymbol{x}) \in \mathbb{R}$ denote affine mappings of $\boldsymbol{x}$ such that $\boldsymbol{a}(\boldsymbol{x}) = \begin{pmatrix}\eta_1 \boldsymbol{x}\\ \eta_2\end{pmatrix}$ and $b_i(\boldsymbol{x}) = \boldsymbol{B}_i^\top \boldsymbol{x} + b^i$ with parameters $\eta_1, \eta_2 \in \{0,1\}, \eta_1 + \eta_2 \ge 1$, $\boldsymbol{B}_i \in \mathbb{R}^n$, and $b^i \in \mathbb{R}$, respectively. For notational convenience, we let $\Xi \subseteq \prod_{i\in[I]} \Xi_i$ and $\tilde{\boldsymbol{\xi}} = (\tilde{\boldsymbol{\xi}}_1, \ldots, \tilde{\boldsymbol{\xi}}_I)$. Note that (i) for any $i, j \in [I]$ and $i \neq j$, the random vectors $\tilde{\boldsymbol{\xi}}_i$ and $\tilde{\boldsymbol{\xi}}_j$ can be correlated; and (ii) we use η_1, η_2 to differentiate whether (1c) involves left-hand uncertainty (i.e., $\eta_1 = 1, \eta_2 = 0$), right-hand uncertainty (i.e., $\eta_1 = 0, \eta_2 = 1$) or both-side uncertainty (i.e., $\eta_1 = 1, \eta_2 = 1$).

The distributionally robust chance constraint (DRCC) (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 $\eta_1 = 0, \eta_2 = 1$ and a DRCCP with *left-hand* uncertainty if $\eta_1 = 1, \eta_2 = 0$. 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 DRCC (1c) as

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

1.2 Assumptions

In this paper, we consider Wasserstein ambiguity set $\mathcal{P}$, i.e., we make the following assumption on the ambiguity set $\mathcal{P}$.

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

$$\mathcal{P}^W = \left\{\mathbb{P} : \mathbb{P}\left\{\tilde{\boldsymbol{\xi}} \in \Xi\right\} = 1, W\left(\mathbb{P}, \mathbb{P}_{\tilde{\zeta}}\right) \le \delta\right\}, \tag{3}$$

where Wasserstein distance is defined as

$$W(\mathbb{P}_1, \mathbb{P}_2) = \inf \left\{ \int_{\Xi \times \Xi} \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\| \mathbb{Q}(d\boldsymbol{\xi}_1, d\boldsymbol{\xi}_2) : \begin{array}{l} \mathbb{Q} \text{ is a joint distribution of } \widehat{\boldsymbol{\xi}}_1 \text{ and } \widehat{\boldsymbol{\xi}}_2 \\ \text{with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2, \text{ respectively} \end{array} \right\},$$

and $\mathbb{P}_{\tilde{\zeta}}$ denotes a discrete empirical distribution of $\tilde{\zeta}$ generated by i.i.d. samples $\mathcal{Z} = \{\zeta^j\}_{j\in[N]} \subseteq \Xi$ from the true distribution $\mathbb{P}^\infty$, i.e., its point mass function is $\mathbb{P}_{\tilde{\zeta}}\left\{\tilde{\zeta} = \zeta^j\right\} = \frac{1}{N}$, and $\delta > 0$ denotes the Wasserstein radius. We assume that $(\Xi, \|\cdot\|)$ is a totally bounded Polish (separable complete metric) space with distance metric $\|\cdot\|$, 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 [12]. The fact that the convergence result is not affected by the support motivates us to consider relaxing the support $\Xi = \mathbb{R}^{I\times(n+1)}$, which provides us better reformulation power. That is, we make the following assumption about the support Ξ.

(A2) The support $\Xi = \mathbb{R}^{I\times(n+1)}$, i.e., $\Xi = \prod_{i\in[I]} \Xi_i$ and $\Xi_i = \mathbb{R}^{n+1}$.

We remark that

- This assumption has been studied in recent DRCCP literature [9, 28, 30];
- By making this assumption, it might cause the DRCC (1c) to be more conservative than the general setting studied in [15];
- The interdependence between different random vectors $\boldsymbol{\xi}_i$ can be inherited implicitly from the empirical distribution. For example, suppose that in the true distribution $\mathbb{P}^\infty$, we have $\mathbb{P}^\infty\left\{\boldsymbol{\xi}_{i_1} = \boldsymbol{\xi}_{i_2}\right\} = 1$ for some $i_1, i_2 \in [I]$ and $i_1 \neq i_2$, then for any empirical sample $j \in [N]$, we must have $\boldsymbol{\zeta}^j_{i_1} = \boldsymbol{\zeta}^j_{i_2}$ with probability one. Since the empirical distribution will converge to the true distribution $\mathbb{P}^\infty$ according to Lemma 3.7 in [11] (i.e., when $N \to \infty$, $\delta \to 0$), thus Wasserstein Ambiguity set (3) will eventually pick up the fact that $\mathbb{P}^\infty\left\{\boldsymbol{\xi}_{i_1} = \boldsymbol{\xi}_{i_2}\right\} = 1$. However, this process might require many more samples than that without Assumption (A2);
- In practice, one needs to choose a proper Wasserstein radius δ through cross validation [11] to alleviate the over-conservatism caused by Assumption (A2).

Finally, we suppose that Assumptions (A1) and (A2) hold throughout the paper.

1.3 Contributions

In this paper, we study approximations and exact reformulations of DRCCP under a 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, i.e.,

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

   where $f(\cdot,\cdot)$ is defined in Theorem 1.

2. We show that set Z, once bounded, is mixed integer representable with big-M coefficients and N additional binary variables.

3. We derive inner and outer approximations based on a **CVaR** interpretation. We develop compact formulations for these approximations and compare their strengths.

4. 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 exploiting the polyhedral properties of submodular functions, we propose a new big-M free mixed integer linear reformulation, which can be effectively solved by a branch and cut algorithm.

2 Exact Reformulations

In this section, we will show that DRCC set Z admits a conditional-value-at-risk (**CVaR**) interpretation and is mixed integer representable. This reformulation also allows us to derive tight inner and outer approximations in the next section.

2.1 CVaR Reformulation

In this subsection, we will reformulate the set Z into its deterministic counterpart with respect to an empirical distribution. The main idea of this reformulation is that we first use the strong duality result from [5, 13] to formulate the worst-case chance constraint into its dual form, and then break down the indicator function according to its definition.

Theorem 1. *Set Z is equivalent to*

$$Z = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{l} \delta - \epsilon\gamma \leq \dfrac{1}{N} \displaystyle\sum_{j\in[N]} \min\left\{ f(\boldsymbol{x},\boldsymbol{\zeta}^j) - \gamma, 0 \right\}, \quad (4a) \\ \gamma \geq 0, \quad (4b) \end{array} \right\}$$

where

$$f(\boldsymbol{x},\boldsymbol{\zeta}) = \min\left\{ \min_{i\in[I]\setminus\mathcal{I}(\boldsymbol{x})} \frac{\max\left\{b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i, 0\right\}}{\|\boldsymbol{a}(\boldsymbol{x})\|_*}, \min_{i\in\mathcal{I}(\boldsymbol{x})} \chi_{\{\boldsymbol{x}: b_i(\boldsymbol{x})<0\}}(\boldsymbol{x}) \right\}, \quad (5)$$

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

Please note that in Theorem 1, we use the fact that $\delta > 0$ from Assumption (A1), and the formulation (4) does not hold if $\delta = 0$.

An interesting corollary of Theorem 1 is that set Z can be reformulated as a conditional-value-at-risk (**CVaR**) constrained set. Before showing this interpretation, let us first introduce 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**) [22] 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 (4) has a **CVaR** interpretation.

Corollary 1. *Set Z is equivalent to*

$$Z = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \frac{\delta}{\epsilon} + \mathbf{CVaR}_{1-\epsilon}\left[-f(\boldsymbol{x}, \tilde{\boldsymbol{\zeta}})\right] \leq 0 \right\}, \tag{6}$$

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

In the following sections, we will derive the inner and outer approximations mainly based upon the **CVaR** formulation in Corollary 1.

2.2 Exact Mixed Integer Program Reformulation

In this subsection, we show that set Z is mixed integer representable. To do so, we first observe that the reformulation of set Z in Theorem 1 can be further simplified as a disjunction of a nonconvex set and a convex set.

Proposition 1. *Set $Z = Z_1 \cup Z_2$, where*

$$Z_1 = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \delta\nu - \epsilon\gamma \leq \frac{1}{N}\sum_{j\in[N]} z_j, & \text{(7a)} \\ z_j + \gamma \leq \max\left\{ b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j, 0 \right\}, \forall i \in [I], j \in [N], & \text{(7b)} \\ z_j \leq 0, \forall j \in [N], & \text{(7c)} \\ \|\boldsymbol{a}(\boldsymbol{x})\|_* \leq \nu, & \text{(7d)} \\ \nu > 0, \gamma \geq 0, & \text{(7e)} \end{array} \right\}$$

and

$$Z_2 = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}(\boldsymbol{x}) = \mathbf{0}, b_i(\boldsymbol{x}) \geq 0, \forall i \in [I] \right\}. \tag{8}$$

We make the following remarks about the disjunctive formulation of set Z.

Remark 1. (i) Set Z_2 is trivial:

- For DRCCP with left-hand uncertainty (i.e., $\eta_1 = 1, \eta_2 = 0$), we have

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

- For DRCCP with right-hand uncertainty or two-side uncertainty (i.e., $\eta_1 \in \{0,1\}, \eta_2 = 1$), we have $Z_2 = \emptyset$.

(ii) According to Lemma 2 in [27], the feasible region induced by a chance constraint is closed, so is set Z. However, set Z_1 might not be closed due to $\nu > 0$ in (7e). In practice, one can find a lower bound $0 < \underline{\nu}$ such that

$$\underline{\nu} \leq \inf_{\boldsymbol{x}\in Z_1} \left\{ \|\boldsymbol{a}(\boldsymbol{x})\|_* : \|\boldsymbol{a}(\boldsymbol{x})\|_* \neq 0 \right\};$$

or let $\underline{\nu}$ be a sufficiently small number. Then replace the constraint $\nu > 0$ in (7e) by $\nu \geq \underline{\nu}$.

We observe that set Z_1 can be formulated as a mixed integer set when it is bounded, i.e., we can use binary variables to represent the nonlinear constraints (7b) as mixed integer linear ones. This result has been observed independently by [8] (see their Proposition 1) for single DRCCP.

Theorem 2. *Suppose there exists an* $\boldsymbol{M} \in \mathbb{R}_+^N$ *such that*

$$\max_{i\in[I]}\max_{\boldsymbol{x}\in Z_1}\left\{\left|b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j\right|\right\} \le M_j$$

for all $j \in [N]$*. Then* Z_1 *is mixed integer representable, i.e.,*

$$Z_1 = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \delta\nu - \epsilon\gamma \le \frac{1}{N}\sum_{j\in[N]} z_j, & \text{(9a)} \\ z_j + \gamma \le s_j, \forall j \in [N], & \text{(9b)} \\ s_j \le b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j + M_j(1-y_j), \forall i \in [I], j \in [N], & \text{(9c)} \\ s_j \le M_j y_j, \forall j \in [N], & \text{(9d)} \\ \|\boldsymbol{a}(\boldsymbol{x})\|_* \le \nu, & \text{(9e)} \\ \nu > 0, \gamma \ge 0, s_j \ge 0, z_j \le 0, y_j \in \{0,1\}, \forall j \in [N]. & \text{(9f)} \end{array} \right\}$$

Usually, we can derive the big-M coefficients by inspection; for example, suppose that $\boldsymbol{x} \in [\boldsymbol{L}, \boldsymbol{U}]$, then for each $j \in [N]$, we can find M_j in the following way: (i) rewrite $b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j = \sum_{\tau\in[n]}(B_{i\tau} - \eta_1\zeta_{i\tau}^j)x_\tau + b^i - \eta_2\zeta_{i(n+1)}^j$ for each $i \in [I]$, (ii) define sets $\widehat{S}_+ = \{\tau \in [n] : B_{i\tau} - \eta_1\zeta_{i\tau}^j > 0\}$ and $\widehat{S}_- = [n] \setminus \widehat{S}_+$, and (iii) let M_j be

$$M_j := \max_{i\in[I]}\max\left\{\sum_{\tau\in\widehat{S}_+}(B_{i\tau} - \eta_1\zeta_{i\tau}^j)U_\tau + \sum_{\tau\in\widehat{S}_-}(B_{i\tau} - \eta_1\zeta_{i\tau}^j)L_\tau + b^i - \eta_2\zeta_{i(n+1)}^j,\right.$$
$$\left. -\sum_{\tau\in\widehat{S}_+}(B_{i\tau} - \eta_1\zeta_{i\tau}^j)L_\tau - \sum_{\tau\in\widehat{S}_-}(B_{i\tau} - \eta_1\zeta_{i\tau}^j)U_\tau - b^i + \eta_2\zeta_{i(n+1)}^j\right\}.$$

There are various methods introduced in literature [21, 23] to further tighten big-M coefficients.

Formulation (9) involves N binary variables and big-M coefficients. In Section 4, we will show that for binary DRCCP, set Z_1 can be reformulated as a big-M free formulation without introducing additional binary variables.

3 Outer and Inner Approximations

In this section, we will introduce one outer approximation and three different inner approximations by exploiting the exact reformulations in the previous section. The outer approximation can provide a lower bound for DRCCP, while inner approximations can provide good-quality feasible solutions.

3.1 VaR Outer Approximation

Note from [22] that for any random variable $\tilde{X}$, we have

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

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. *Set Z can be outer approximated by*

$$Z_{\mathbf{VaR}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \mathbb{P}_{\tilde{\zeta}} \left\{ \frac{\delta}{\epsilon} \|\boldsymbol{a}(\boldsymbol{x})\|_* + \boldsymbol{a}(\boldsymbol{x})^\top \tilde{\boldsymbol{\zeta}}_i \le b_i(\boldsymbol{x}), i \in [I] \right\} \ge 1 - \epsilon \right\}. \tag{10}$$

We make the following remarks about outer approximation $Z_{\mathbf{VaR}}$.

(i) In (10), 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, 18]);

(ii) A particular interpretation of formulation (10) is that in order to enforce the robustness, we further penalize the left-hand side of uncertain constraints by the dual norm $\|\boldsymbol{a}(\boldsymbol{x})\|_*$; and

(iii) Suppose that the empirical distribution will converge to the true distribution $\mathbb{P}^\infty$ (cf., Lemma 3.7 in [11]), i.e., $\delta \to 0$ as $N \to \infty$. Then $Z_{\mathbf{VaR}} \to Z$ as $N \to \infty$.

This final remark is summarized below.

Proposition 2. *Suppose that the empirical distribution $\mathbb{P}_{\tilde{\zeta}}$ will converge to the true distribution $\mathbb{P}^\infty$. Then with probability one, we have $Z_{\mathbf{VaR}} \to Z$ as $N \to \infty$.*

Recently, there are several works [3, 4, 14, 25] on distributionally robust optimization with $\infty-$Wasserstein ambiguity set, and set $Z_{\mathbf{VaR}}$ is in fact equal to the feasible region induced by DRCC with $\infty-$Wasserstein ambiguity set.

Proposition 3. *Consider $\infty-$Wasserstein ambiguity set $\mathcal{P}^W$ defined as*

$$\mathcal{P}^W_\infty = \left\{ \mathbb{P} : \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} \in \Xi \right\} = 1, W_\infty\left(\mathbb{P}, \mathbb{P}_{\tilde{\zeta}}\right) \le \frac{\delta}{\epsilon} \right\}, \tag{11}$$

where $\infty-$Wasserstein distance is defined as

$$W_\infty(\mathbb{P}_1, \mathbb{P}_2) = \inf_{\mathbb{Q}} \left\{ \text{ess.sup}\|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\| : \begin{array}{l} \mathbb{Q} \text{ is a joint distribution of } \widehat{\boldsymbol{\xi}}_1 \text{ and } \widehat{\boldsymbol{\xi}}_2 \\ \text{with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2, \text{ respectively} \end{array} \right\}.$$

Then set $Z_{\mathbf{VaR}}$ is equivalent to

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

This result demonstrates that set $Z_{\mathbf{VaR}}$ indeed can be viewed as a deterministic counterpart of DRCCP with $\infty-$Wasserstein ambiguity set. Thus, in practice, it can serve as an alternative for the set Z.

3.2 Inner Approximation I- Robust Scenario Approximation

We also observe that for any random variable $\tilde{X}$, we have

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

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. *Set Z can be inner approximated by*

$$Z_R = \left\{ \mathbf{x} \in \mathbb{R}^n : \frac{\delta}{\epsilon} \|\mathbf{a}(\mathbf{x})\|_* + \mathbf{a}(\mathbf{x})^\top \boldsymbol{\zeta}^j \leq b_i(\mathbf{x}), \forall j \in [N], i \in [I] \right\}. \tag{12}$$

We remark that set Z_R in (12) is very similar to scenario approach to regular chance constrained program [6, 7, 20]. That is, we generate N i.i.d. samples $\{\boldsymbol{\zeta}^j\}_{j\in[N]}$ and enforce all the sampled constraints to hold. It has been shown in [6, 7, 20] that if N is larger than a threshold, it guarantees with high probability that the solution of scenario approach is feasible to the regular chance constrained program. Different from scenario approach, in formulation (12), we add a penalty $\frac{\delta}{\epsilon}\|\mathbf{a}(\mathbf{x})\|_*$ to the sampled constraints, which can be viewed as a "robust" scenario approach to the regular chance constrained problem. That is, if the sample size N is not sufficiently large (i.e., N is smaller than the threshold given by [6, 7, 20]), one might want to add a penalty $\frac{\delta}{\epsilon}\|\mathbf{a}(\mathbf{x})\|_*$ to enforce that set Z_R is indeed a subset of the feasible region induced by a regular chance constraint.

3.3 Inner Approximation II- An Inner Chance Constrained Programming Approximation

Next we propose an inner chance constrained programming approximation of set Z by constructing a feasible γ in (4).

Theorem 5. *Set Z is inner approximated by*

$$Z_I = \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbb{P}_{\tilde{\zeta}} \left\{ \frac{\delta}{\epsilon - \alpha} \|\mathbf{a}(\mathbf{x})\|_* + \mathbf{a}(\mathbf{x})^\top \tilde{\boldsymbol{\zeta}}_i \leq b_i(\mathbf{x}), i \in [I] \right\} \geq 1 - \alpha, 0 \leq \alpha < \epsilon \right\}. \tag{13}$$

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

We also observe that (i) set Z_R is a special case of set Z_I by letting $\alpha = 0$, thus, we must have $Z_R \subseteq Z_I$; (ii) there are $\lceil N\epsilon \rceil$ disjoint intervals that α belongs to, that is,

$$\alpha \in \cup_{i\in[\lceil N\epsilon \rceil]} \left[\frac{i-1}{N}, \frac{i}{N} \right).$$

Suppose that $\alpha \in (\frac{i-1}{N}, \frac{i}{N})$ for some $i \in [\lceil N\epsilon \rceil]$. Since $\mathbb{P}_{\tilde{\zeta}} \{\tilde{\boldsymbol{\zeta}} = \boldsymbol{\zeta}^j\} = \frac{1}{N}$ for all $j \in [N]$, thus the chance constraint in (13) is equivalent to

$$\mathbb{P}_{\tilde{\zeta}} \left\{ f(\mathbf{x}, \tilde{\boldsymbol{\zeta}}) \geq \frac{\delta}{\epsilon - \alpha} \right\} \geq 1 - \frac{i-1}{N}.$$

The feasible region induced by the above chance constraint increases if we decrease the value of α to $\frac{i-1}{N}$. Therefore, to optimize over set $S \cap Z_I$, we only need to enumerate these $\lceil N\epsilon \rceil$ different values of α and choose the one which yields the smallest objective value; (iii) for each given α, the chance constraint in (13) is mixed integer program representable.

Finally, suppose that the empirical distribution $\mathbb{P}_{\tilde{\zeta}}$ will converge to the true distribution $\mathbb{P}^\infty$ with an exponential rate (cf., Theorem 3.4 in [11]), i.e., if $N \to \infty$, then $\delta \to 0$ with rate $\delta = \frac{c_1}{N^{c_2}}$, where $c_1 > 0, c_2 > 0$ are positive constants. Then with probability one, $Z_I \to Z$ as $N \to \infty$. Indeed, suppose that N is sufficiently large such that $\frac{c_1}{N^{\frac{c_2}{2}}} < 1$. In Z_I, let $\alpha = \frac{\lceil N\epsilon \rceil - \lceil c_1 N^{1-\frac{c_2}{2}} \rceil - 1}{N}$. Clearly, as $N \to \infty$, we have $\alpha \to \epsilon$ and

$$\frac{\delta}{\epsilon - \alpha} = \frac{c_1 N^{1-c_2}}{N\epsilon + 1 - \lceil N\epsilon \rceil + \lceil c_1 N^{1-\frac{c_2}{2}} \rceil} \leq N^{-\frac{c_2}{2}} \to 0$$

where the inequality is due to $N\epsilon + 1 \geq \lceil N\epsilon \rceil$ and $\lceil c_1 N^{1-\frac{c_2}{2}} \rceil \geq c_1 N^{1-\frac{c_2}{2}}$. This observation is summarized below.

Proposition 4. *Suppose that the empirical distribution $\mathbb{P}_{\tilde{\zeta}}$ will converge to the true distribution $\mathbb{P}^{\infty}$ with an exponential rate. Then with probability one, we have $Z_I \to Z$ as $N \to \infty$.*

We make the following two remarks:

- According to [11], any light-tail distribution (e.g., a Gaussian distribution) satisfies the assumption in the above proposition; and
- Sets $Z_{\mathbf{VaR}}$ and Z_I together build up a hierarchy of regular chance constrained programs, which converges to DRCC set Z as $N \to \infty$ and preserves the outer and inner approximations, i.e., $Z_I \subseteq Z \subseteq Z_{\mathbf{VaR}}$ for all N and $Z_I \to Z$ and $Z_{\mathbf{VaR}} \to Z$ as $N \to \infty$.

3.4 Inner Approximation III- CVaR Approximation

In this subsection, we will study 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., [19]). For DRCC set Z, the resulting approximation is

$$Z_{\mathbf{CVaR}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \sup_{\mathbb{P} \in \mathcal{P}} \inf_{\beta} \left[-\epsilon\beta + \mathbb{E}_{\mathbb{P}} \left[\left(\max_{i \in [I]} \left(\boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi}_i - b_i(\boldsymbol{x}) \right) + \beta \right)_+ \right] \right] \leq 0 \right\}. \tag{14}$$

Set $Z_{\mathbf{CVaR}}$ (14) 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 [16]. Thus, the proof is omitted.

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

$$Z_{\mathbf{CVaR}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \begin{array}{ll} \delta\nu - \epsilon\gamma \leq \frac{1}{N} \sum_{j \in [N]} z_j, & \text{(15a)} \\ z_j + \gamma \leq b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j, \forall j \in [N], i \in [I], & \text{(15b)} \\ z_j \leq 0, \forall j \in [N], & \text{(15c)} \\ \|\boldsymbol{a}(\boldsymbol{x})\|_* \leq \nu, & \text{(15d)} \\ \nu \geq 0, \gamma \geq 0. & \text{(15e)} \end{array} \right\}$$

We can show that if $N\epsilon \leq 1$, then we must have $f(\boldsymbol{x}, \boldsymbol{\zeta}^j) = \underline{f}(\boldsymbol{x}, \boldsymbol{\zeta}^j)$ for all $j \in [N]$, which implies that $Z_{\mathbf{CVaR}} = Z$.

Proposition 5. *Suppose that $\epsilon \in (0, 1/N]$, then $Z = Z_{\mathbf{CVaR}}$.*

The result in Proposition 5 shows that if the risk parameter ϵ is small enough (i.e., less than or equal to $\frac{1}{N}$), then set Z is convex and is equivalent to its **CVaR** approximation.

3.5 Formulation Comparisons

First, we would like to compare sets $Z_R, Z_{\mathbf{CVaR}}$. Indeed, we can show that $Z_R \subseteq Z_{\mathbf{CVaR}}$, i.e., set Z_R is at least as conservative as **CVaR** approximation $Z_{\mathbf{CVaR}}$.

Proposition 6. *Let* $Z_R, Z_{\mathbf{CVaR}}$ *be defined in* (12), (15) *, respectively. Then* $Z_R \subseteq Z_{\mathbf{CVaR}}$.

The following example illustrates sets $Z, Z_{\mathbf{VaR}}, Z_{\mathbf{CVaR}}, Z_R, Z_I$ and their inclusive relationships.

Example 1. Suppose $N = 3, n = 2, I = 2, \delta = 1/6, \epsilon = 2/3$ and $\zeta_1^1 = (0, 0, \sqrt{2})^\top, \zeta_2^1 = (0, 0, 3\sqrt{2})^\top, \zeta_1^2 = (0, 0, 3\sqrt{2})^\top, \zeta_2^2 = (0, 0, \sqrt{2})^\top, \zeta_1^3 = (0, 0, 3\sqrt{2})^\top, \zeta_2^3 = (0, 0, 2\sqrt{2})^\top, \boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{e}_3 = (0, 0, 1)^\top, b_1(x) = x_1, b_2(x) = x_2$. Then, (2) becomes:

$$Z := \left\{ (x_1, x_2) : \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left\{ (\tilde{\xi}_1, \tilde{\xi}_2) : \tilde{\xi}_{13} \le x_1, \tilde{\xi}_{23} \le x_2 \right\} \ge \frac{1}{3} \right\}. \tag{16}$$

By straightforward calculation, we have $Z_R \subsetneq \left\{ \begin{smallmatrix} Z_{\mathbf{CVaR}} \\ \not\supseteq\!\!\not\subseteq \\ Z_I \end{smallmatrix} \right\} \subsetneq Z \subsetneq Z_{\mathbf{VaR}}$ (see Figure 1 for an illustration).

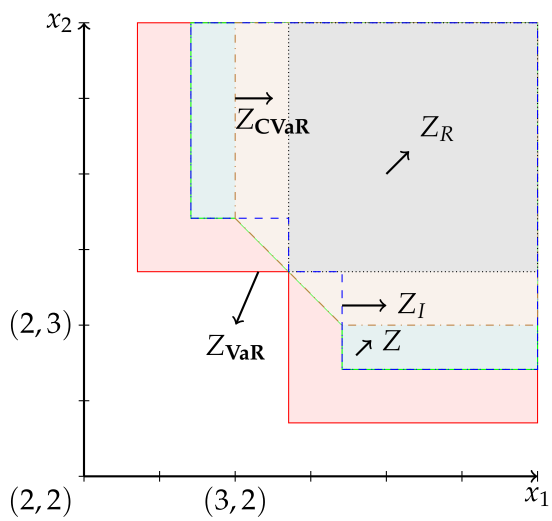

Figure 1: Illustration of Example 1

Finally, the theoretical inclusive relationships of sets $Z, Z_{\mathbf{VaR}}, Z_R, Z_I, Z_{\mathbf{CVaR}}$ are shown in Figure 2.

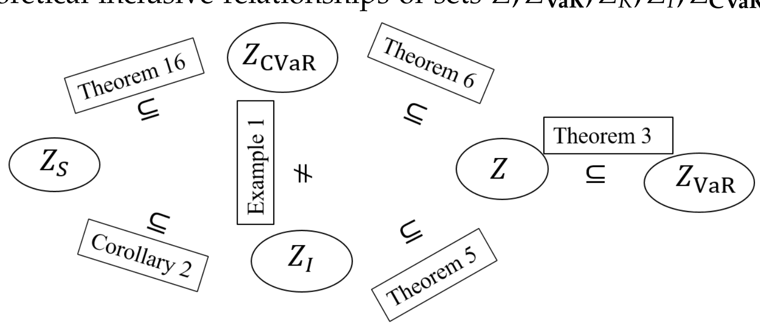

Figure 2: Summary of formulation comparisons

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., we assume that $S \subseteq \{0,1\}^n$. If S is a bounded integer set, we can use binary expansion to reformulate S as an equivalent binary set (c.f., [31]). 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 briefly introduce the definition of submodularity and interested readers are referred to [10, 17] for more details.

Definition 1. **(Submodularity)** Let $2^{[n]}$ be the power set of $[n]$. Then a set function $g : 2^{[n]} \to \mathbb{R}$ is "*submodular*" if and only if it satisfies the following condition:

- for every $T_1, T_2 \subseteq [n]$ with $T_1 \subseteq T_2$ and every $t \in [n] \setminus T_2$, we must have $g(T_1 \cup \{t\}) - g(T_1) \geq g(T_2 \cup \{t\}) - g(T_2)$.

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. Since $\boldsymbol{d}_1^\top \boldsymbol{x} + d_2$ is a nondecreasing submodular function and $-\max(t, d_3)$ is a nonincreasing concave function, the submodularity of their composition follows by Table 1 in [24]. □ □

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

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, 29], i.e.,

$$\operatorname{conv}\left(\Pi_f\right) = \left\{(\boldsymbol{x}, \phi) : f(\boldsymbol{0}) + \sum_{l \in [n]} \rho_{\sigma_l} x_{\sigma_l} \leq \phi, \forall \sigma \in \Omega, \boldsymbol{x} \in [0,1]^n\right\}, \tag{17}$$

where Ω denotes a collection of all 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 (17), these inequalities can be easily separated by a greedy procedure.

Lemma 3. *([2, 29]) Suppose $(\tilde{\boldsymbol{x}}, \tilde{\phi}) \notin \operatorname{conv}\left(\Pi_f\right)$, and let $\sigma \in \Omega$ 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 $\operatorname{conv}\left(\Pi_f\right)$, 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 $\operatorname{conv}\left(\Pi_f\right)$. 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 nonlinear constraints defining the feasible region of a binary DRCCP (i.e., set $S \cap Z$) by submodular knapsack constraints. These constraints can be equivalently described by the system of EPI in (17). Therefore we obtain a big-M free mixed integer representation of set $S \cap Z$.

First, we introduce n auxiliary variables complementing binary variables $\boldsymbol{x}$, 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

$$b_i(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\zeta}_i^j = \boldsymbol{r}_{ij}^\top \boldsymbol{x} + \boldsymbol{t}_{ij}^\top \boldsymbol{w} + u_{ij} \tag{18}$$

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 $\boldsymbol{a}(\boldsymbol{x}) = \begin{pmatrix} \eta_1 \boldsymbol{x} \\ \eta_2 \end{pmatrix}$ and $b_i(\boldsymbol{x}) = \boldsymbol{B}_i^\top \boldsymbol{x} + b^i$, in (18), we can choose

$$\begin{aligned}
r_{ijl} &= B_{il}\mathbb{I}(B_{il} > 0) - \eta_1 \zeta_{il}^j \mathbb{I}(\zeta_{il}^j < 0), \\
t_{ijl} &= -B_{il}\mathbb{I}(B_{il} < 0) + \eta_1 \zeta_{il}^j \mathbb{I}(\zeta_{il}^j > 0), \\
u_{ij} &= b^i - \eta_2 \boldsymbol{\zeta}_{i(n+1)}^j + \sum_{\tau \in [n]} \left(B_{i\tau}\mathbb{I}(B_{i\tau} < 0) - \eta_1 \zeta_{i\tau}^j \mathbb{I}(\zeta_{i\tau}^j > 0) \right),
\end{aligned}$$

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

Thus, from above discussion, we can formulate $S \cap Z$ (recall that set $Z = Z_1 \cup Z_2$ according to Proposition 1) as the following mixed integer set with submodular knapsack constraints.

Theorem 7. *Suppose that $S \subseteq \{0,1\}^n$. Then $S \cap Z = (S \cap \widehat{Z}_1) \cup (S \cap Z_2)$, where*

$$S \cap \widehat{Z}_1 = \left\{ \boldsymbol{x} \in S : \begin{array}{ll}
\delta \nu - \epsilon \gamma \le \frac{1}{N} \sum_{j \in [N]} z_j, & \text{(19a)} \\
-\max\left\{ \boldsymbol{r}_{ij}^\top \boldsymbol{x} + \boldsymbol{t}_{ij}^\top \boldsymbol{w} + u_{ij}, 0 \right\} \le -z_j - \gamma, \forall i \in [I], j \in [N], & \text{(19b)} \\
z_j \le 0, \forall j \in [N], & \text{(19c)} \\
\left\| \begin{pmatrix} \eta_1 \boldsymbol{x} \\ \eta_2 \end{pmatrix} \right\|_* \le \nu, & \text{(19d)} \\
w_l + x_l = 1, \forall l \in [n], & \text{(19e)} \\
\nu \ge 1, & \text{(19f)} \\
\gamma \ge 0, \boldsymbol{w} \in \{0,1\}^n & \text{(19g)}
\end{array} \right\}$$

and

$$S \cap Z_2 = \{ \boldsymbol{x} \in S : \boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{0}, b_i(\boldsymbol{x}) \ge 0, \forall i \in [I] \} \tag{20}$$

We note that the left-hand sides of constraints (19b) and (19d) are submodular functions according to Lemma 1 and Lemma 2. Therefore, equivalently, we can replace these constraints with the convex hulls of epigraphs of their associated submodular functions. Thus, we arrive at the following equivalent representation of set $S \cap \widehat{Z}_1$.

Corollary 2. *Suppose that $S \subseteq \{0,1\}^n$ and $\|\cdot\|$ is L_p norm with $p \geq 1$. Then*

$$S \cap \widehat{Z}_1 = \left\{ \boldsymbol{x} \in S : \begin{array}{ll} \delta\nu - \epsilon\gamma \leq \frac{1}{N}\sum_{j\in[N]} z_j, & \text{(21a)} \\ (\boldsymbol{x}, \boldsymbol{w}, -z_j - \gamma) \in \mathrm{conv}(\Pi_{ij}), \forall i \in [I], j \in [N], & \text{(21b)} \\ z_j \leq 0, \forall j \in [N], & \text{(21c)} \\ (\boldsymbol{x}, \nu) \in \mathrm{conv}(\Pi_0), & \text{(21d)} \\ w_l + x_l = 1, \forall l \in [n], & \text{(21e)} \\ \nu \geq 1, \gamma \geq 0, \boldsymbol{w} \in [0,1]^n, & \text{(21f)} \end{array} \right\}$$

where

$$\Pi_{ij} = \left\{ (\boldsymbol{x}, \boldsymbol{w}, \phi) : -\max\left\{ \boldsymbol{r}_{ij}^\top \boldsymbol{x} + \boldsymbol{t}_{ij}^\top \boldsymbol{w} + u_{ij}, 0 \right\} \leq \phi, \boldsymbol{x}, \boldsymbol{w} \in \{0,1\}^n \right\}, \forall i \in [I], j \in [N], \quad \text{(22a)}$$

$$\Pi_0 = \left\{ (\boldsymbol{x}, \phi) : \left\| \begin{pmatrix} \eta_1 \boldsymbol{x} \\ \eta_2 \end{pmatrix} \right\|_* \leq \phi, \boldsymbol{x} \in \{0,1\}^n \right\} \quad \text{(22b)}$$

and $\left\{\mathrm{conv}(\Pi_{ij})\right\}_{i\in[I], j\in[N]}, \mathrm{conv}(\Pi_0)$ *can be described by the system of EPI in* (17).

Note that the optimization problem $\min_{\boldsymbol{x}\in S\cap Z_1} \boldsymbol{c}^\top \boldsymbol{x}$ can be solved by a 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 (21b) and (21d), there are $N \times I + 1$ convex hulls of epigraphs (i.e., $\left\{\mathrm{conv}(\Pi_{ij})\right\}_{i\in[I], j\in[N]}, \mathrm{conv}(\Pi_0)$) to be separated from. Therefore, instead, we can first check and find the epigraphs of κ (e.g., $\kappa = 10$ in our numerical study) most violated constraints in (19b) and (19d), i.e., find the epigraphs corresponding to the κ largest values in the following set

$$\left\{ -\max\left\{ \boldsymbol{r}_{ij}^\top \widehat{\boldsymbol{x}} + \boldsymbol{t}_{ij}^\top \widehat{\boldsymbol{w}} + u_{ij}, 0 \right\} + \widehat{z}_j + \widehat{\gamma} \right\}_{i\in[I], j\in[N]} \bigcup \left\{ \left\| \begin{pmatrix} \eta_1 \widehat{\boldsymbol{x}} \\ \eta_2 \end{pmatrix} \right\|_* - \widehat{\nu} \right\}.$$

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.

5 Conclusion

In this paper, we studied a distributionally robust chance constrained problem (DRCCP) with Wasserstein ambiguity set. We showed that a DRCCP could be formulated as a conditional value-at-risk constrained optimization problem, and thus admits tight inner and outer approximations. If the feasible region is bounded, we showed that a DRCCP could be 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 and a branch and cut solution algorithm.

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] Dimitris Bertsimas, Shimrit Shtern, and Bradley Sturt. A data-driven approach for multi-stage linear optimization. Available at Optimization Online, 2018.

[4] Dimitris Bertsimas, Shimrit Shtern, and Bradley Sturt. Two-stage sample robust optimization. *arXiv preprint arXiv:1907.07142*, 2019.

[5] Jose Blanchet and Karthyek RA Murthy. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research.*, 2018.

[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] 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.

[8] Zhi Chen, Daniel Kuhn, and Wolfram Wiesemann. Data-driven chance constrained programs over Wasserstein balls. *arXiv preprint arXiv:1809.00210*, 2018.

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

[10] Jack Edmonds. Submodular functions, matroids, and certain polyhedra. *Combinatorial structures and their applications*, pages 69–87, 1970.

[11] Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166, 2018.

[12] 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.

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

[14] Rui Gao and Anton J Kleywegt. Distributionally robust stochastic optimization with dependence structure. *arXiv preprint arXiv:1701.04200*, 2017.

[15] 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.

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

[17] László Lovász. Submodular functions and convexity. In *Mathematical Programming The State of the Art*, pages 235–257. Springer, 1983.

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

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

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

[21] 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.

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

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

[24] Donald M Topkis. Minimizing a submodular function on a lattice. *Operations research*, 26(2):305–321, 1978.

[25] Weijun Xie. Tractable reformulations of distributionally robust two-stage stochastic programs with $\infty-$Wasserstein distance. Available at Optimization Online, 2018.

[26] Weijun Xie. On distributionally robust chance constrained programs with Wasserstein distance. *Mathematical Programming*, pages 1–41, 2019.

[27] 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.

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

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

[30] Yiling Zhang, Ruiwei Jiang, and Siqian Shen. Ambiguous chance-constrained binary programs under mean-covariance information. *SIAM Journal on Optimization*, 28(4):2922–2944, 2018.

[31] Jikai Zou, Shabbir Ahmed, and Xu Andy Sun. Stochastic dual dynamic integer programming. *Mathematical Programming*, 175(1):461–502, 2019.