Distributionally robust joint chance-constrained programming with Wasserstein metric

Yining Gu and Yanjun Wang

School of Mathematics, Shanghai University of Finance and Economics, Shanghai, People’s Republic of China

ABSTRACT
In this paper, we develop an exact reformulation and a deterministic approximation for distributionally robust joint chance-constrained programming (DRCCPs) with a general class of convex uncertain constraints under data-driven Wasserstein ambiguity sets. It is known that robust chance constraints can be conservatively approximated by worst-case conditional value-at-risk (CVaR) constraints. It is shown that the proposed worst-case CVaR approximation model can be reformulated as an optimization problem involving biconvex constraints for joint DRCP. This approximation is essentially exact under certain conditions. We derive a convex relaxation of this approximation model by constructing new decision variables which allows us to eliminate biconvex terms. Specifically, when the constraint function is affine in both the decision variable and the uncertainty, the resulting approximation model is equivalent to a tractable mixed-integer convex reformulation for joint binary DRCP. Numerical results illustrate the computational effectiveness and superiority of the proposed formulations.

1. Introduction
1.1. Problem setting
In this paper, we study distributionally robust chance-constrained programing (DRCCPs) of the form:

\[
\begin{align*}
\min_x & \quad c^T x \\
\text{s.t.} & \quad x \in S, \\
& \quad \inf_{P \in \mathcal{P}} \mathbb{P}\left\{ \hat{\xi} : f_t(x, \hat{\xi}) \geq 0, \forall t \in [T] \right\} \geq 1 - \epsilon, 
\end{align*}
\]

where \( \mathbf{x} \in \mathbb{R}^n \) is a decision vector; the vector \( \mathbf{c} \in \mathbb{R}^n \) represents the objective function coefficients; the set \( S \subseteq \mathbb{R}^n \) represents a computable bounded convex set (e.g. a polyhedron); Let \( (\Omega, \mathcal{F}) \) be a measurable space with \( \mathcal{F} \) being the Borel \( \sigma \)-algebra on \( \Omega \). The random vector \( \hat{\xi} \in \mathbb{R}^m \) is defined on a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) with a closed support set \( \mathcal{X} \).
and the set function $\mathbb{P}\{\cdot\}$ represents the associated probability measure of $\tilde{\xi}$; $\mathcal{P}(\Xi)$ represents the set of all probability measures defined on $\Xi$, and $\mathcal{P} \subset \mathcal{P}(\Xi)$ is termed as an ‘ambiguity set’ comprising all distributions that are compatible with the decision maker’s prior information; the set $F(x) = \{\xi : f_i(x, \xi) \geq 0, \forall t \in [T]\}$ represents the feasible region described by a set of uncertain constraints, in which the mapping $f_i(x, \xi) : \mathbb{R}^n \times \Xi \rightarrow \mathbb{R}$ for any $t \in [T] := \{1, 2, \ldots, T\}$ represents the robust constraint function on $x$. In addition, the distributionally robust chance constraint (DRCC) $(1c)$ requires all $T$ uncertain constraints to be jointly satisfied for all the probability distributions from the ambiguity set $\mathcal{P}$, which can be modelled as bilinear, i.e., $f_i(x, \xi)$ is convex in $\xi$ for any fixed $x$, and is concave in $x$ for any fixed $\xi$.

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

$$Z_D := \left\{ x \in \mathbb{R}^n : \inf_{P \in \mathcal{P}} \mathbb{P}\left\{ \xi : f_i(x, \xi) \geq 0, \forall t \in [T]\right\} \geq 1 - \epsilon \right\}$$

$$= \left\{ x \in \mathbb{R}^n : \sup_{P \in \mathcal{P}} \mathbb{P}\left\{ \tilde{\xi} : f_i(x, \tilde{\xi}) < 0, \exists t \in [T]\right\} \leq \epsilon \right\}$$

$$= \left\{ x \in \mathbb{R}^n : \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[ \mathbb{I}\{f_i(x, \tilde{\xi}) < 0, \exists t \in [T]\}(\tilde{\xi})\right] \leq \epsilon \right\}.$$

Note that all results in the remainder of this paper are predicated on the following assumptions.

(A1): Each function $f_i(x, \xi)$ is convex in $\xi$ for any fixed $x$, and is concave in $x$ for any fixed $\xi$.

(A2): The random vector $\tilde{\xi}$ is supported on a nonempty closed convex set $\Xi \subseteq \mathbb{R}^m$.

We remark that previous research focuses on the special case where $f_i(x, \xi)$ is bilinear, i.e., $f_i(x, \xi)$ is affine in both the decision variable $x$ and the uncertainty $\xi$, and the distribution of $\tilde{\xi}$ ranges over the ambiguity set $\mathcal{P}$ without support constraints. However, the uncertainty is inherently non-linear in many application problems, and cannot be modelled by a bilinear constraint function, e.g. [1,14,31]. Assumptions (A1–A2) imply that this paper is devoted to analyzing the tractability of DRCCP in a more general setting where $f_i(x, \xi)$ is non-linear, i.e., $f_i(x, \xi)$ is convex in the uncertainty $\xi$, and is concave in the decision variable $x$. Indeed, assumption (A1) allows us to consider relaxing the support set $\Xi = \mathbb{R}^m$. But, it should be realized that disregarding the support information of the true distribution $\mathbb{P}_{\text{true}}$ of $\tilde{\xi}$ can result in unnecessarily conservative robust chance constraints. Thus, the support set $\Xi$ of $\tilde{\xi}$ is defined as a nonempty closed convex set. If $\tilde{\xi}$ is known to be supported on a non-convex subset $\Xi$ of $\mathbb{R}^m$, then we can replace $\Xi$ with its convex hull to satisfy assumption (A2). These conditions above provide us better reformulation power.

Now we discuss how these assumptions above are satisfied by the following example.

Example 1.1: Consider a portfolio optimization problem investigated in [14,31]. Assume that the market consists of $n \leq m$ basic assets and $m - n$ derivatives which are identified as European-style call options derived from the basic assets. We denote the present as time
\( t = 0 \) and the end of the investment horizon as \( t = T \). A portfolio is characterized by a vector of asset weights \( w \in \mathbb{R}^m \), whose elements add up to 1. Further we partition the allocation vector as \( w = (w^1, w^2) \), where \( w^1 \in \mathbb{R}^n \) and \( w^2 \in \mathbb{R}^{m-n} \) denote the percentage allocations in the basic assets and options, respectively. Let asset \( j \) be a call option with strike price \( k_j \) on the basic asset \( i \), and denote the initial prices of the option and asset \( i \) by \( c_j \) and \( s_i \), respectively. Moreover, denote the random return of the basic asset \( i \) by \( \tilde{r}_i \), the return of this option \( j \) is \( \frac{1}{c_j} \max\{0, s_i(1 + \tilde{r}_i) - k_j\} - 1 \). Our goal is to find an allocation vector \( w \) that entails a high portfolio return while keeping the associated risk at an acceptable level, which leads to the following formulation by Value-at-Risk (VaR) framework.

\[
\begin{align*}
\min_{w \geq 0, z} & \quad z \\
\text{s.t.} & \quad w^\top e_m = 1, \\
& \quad \mathbb{P} \left\{ z + \sum_{i=1}^n w_i^1 \tilde{r}_i + \sum_{j=1}^{m-n} w_j^2 \max \left\{ 0, \frac{s_i}{c_j} (1 + \tilde{r}_i) - \frac{k_j}{c_j} \right\} - 1 \geq 0 \right\} \geq 1 - \epsilon,
\end{align*}
\]

where \( \mathbb{P} \) denotes the distribution of the basic asset returns \( \tilde{r} \) and \( \epsilon \in (0, 1) \) is a specified risk level. Note that this formulation satisfies assumption (A1), i.e. the constraint function is convex in the uncertainty \( \tilde{r} \), and is concave (linear) in the decision variable \((z, w)\).

In this paper, the ambiguity set employed in the distributionally robust formulation is the Wasserstein ball centred at the empirical distribution of the sample dataset. Moreover, we make the following assumption on the ambiguity set \( \mathcal{P} \).

\textbf{(A3)} The Wasserstein ambiguity set \( \mathcal{P}_W \) is defined as

\[
\mathcal{P}_W = \left\{ \mathbb{P} : \mathbb{P} \left\{ \tilde{\xi} \in \Xi \right\} = 1, W(\mathbb{P}, \mathbb{P}_{\tilde{\xi}}) \leq \delta \right\},
\]

where the 1-Wasserstein metric is defined as

\[
W(\mathbb{P}_1, \mathbb{P}_2) = \inf \left\{ \int_{\Xi \times \Xi} \| \tilde{\xi}_1 - \tilde{\xi}_2 \| \mathbb{Q} (d\tilde{\xi}_1, d\tilde{\xi}_2) : \mathbb{Q} \text{ is a joint distribution of } \tilde{\xi}_1 \text{ and } \tilde{\xi}_2 \text{ with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2, \text{ respectively} \right\}
\]

for all distributions \( \mathbb{P}_1, \mathbb{P}_2 \in \mathcal{M}(\Xi) \), where \( \mathcal{M}(\Xi) \) contains all probability distributions \( \mathbb{P} \) supported on \( \Xi \) with \( \mathbb{E}_\mathbb{P} \| \tilde{\xi} \| = \int_{\Xi} \| \tilde{\xi} \| \mathbb{P}(d\tilde{\xi}) < \infty \).

The 1-Wasserstein metric between \( \mathbb{P}_1 \) and \( \mathbb{P}_2 \), equipped with an arbitrary norm \( \| \cdot \| \) on \( \mathbb{R}^m \), represents the minimum transportation cost generated by moving the probability mass from \( \mathbb{P}_1 \) to \( \mathbb{P}_2 \). In (5), \( \mathbb{P}_{\tilde{\xi}} \) represents a discrete empirical distribution of \( \tilde{\xi} \) with i.i.d. samples \( Z = \{\tilde{\xi}^i\}_{i \in [N]} \subseteq \Xi \) from the true distribution \( \mathbb{P}^\infty \), i.e. its point mass function is \( \mathbb{P}_{\tilde{\xi}}(\tilde{\xi} = \tilde{\xi}^i) = \frac{1}{N} \), and \( \delta > 0 \) represents the Wasserstein radius.

\subsection*{1.2. Literature review}

There are significant efforts on reformulations, approximations and convexity properties of DRCCP problems under various ambiguity sets. The choice of the ambiguity set \( \mathcal{P} \) is
essential in DRCCP problems. It should be large enough to include the true probability distribution with a high confidence but cannot be too large to avoid very conservative decisions. In particular, many approaches based on moments and statistical distances are commonly used to build ambiguity sets in DRCCP problems.

We now review existing works on DRCCP problems with moment-based ambiguity sets [2,7–9,28–30,32]. It is well-known that more efforts have been made to derive tractable reformulations for single DRCCP with constraint functions that are affine in both the decision variable and the uncertainty. For instance, the authors in [2] demonstrated that with given first- and second-order moments, the set $Z_D$ for single DRCCP is equivalent to a tractable second-order conic representation. In [32], the authors developed tractable semidefinite programming for single DRCCP with given first- and second-order moments as well as the support of the uncertain parameters. In addition, the authors in [8] showed that the set $Z_D$ for single DRCCP is convex when $P$ involves conic moment constraints or unimodality of $P$. However, tractability results for joint DRCCP with affine uncertain constraints are very rare. It has been shown in [8] that this kind of joint DRCCP with a richer class of ambiguity sets defined through moment conditions is NP-hard in general. Thus, much of the earlier works derived deterministic approximations of the set $Z_D$ instead of developing its equivalent reformulations. For instance, in [32], with given first- and second-order moments, the authors derived a tractable conservative approximation for joint DRCCP in terms of a worst-case CVaR constraint, which is tighter than the well-known Bonferroni approximation. Besides, the authors in [29] provided several sufficient conditions under which the Bonferroni approximation of joint DRCCP is exact and obtained its convex reformulation when $P$ is specified by first- and second-order moments. On the other hand, there is very limited literature on DRCCP problems with a broader family of uncertain constraints. For instance, in [32], with given first- and second-order moments, the authors proved that single DRCCP amounts to a tractable semidefinite programming when the constraint function is either concave piecewise or (possibly non-concave) quadratic in the uncertainty. With any ambiguity set including convex moment constraints, the authors in [28] studied deterministic reformulations of the set $Z_D$ and its convexity properties when the constraint function is convex in the uncertainty and is concave in the decision variable. They investigated deterministic reformulations of such problems and proposed some conditions under which such deterministic reformulations are convex. In [30], the authors proved that with given first- and second-order moments, single DRCCP is equivalent to a robust optimization problem when the constraint function is quasi-convex in the uncertainty and is concave in the decision variable. With a different perspective from the above literature, the authors in [7] elaborated an approximation approach for joint DRCCP via a sequence of approximated ambiguity sets $\{P_N\}$ and established convergence results for this method. Specifically, with the ambiguity set $P$ being defined through some general moment conditions, they developed a tractable approximation for joint DRCCP by constructing a piecewise uniform approximation scheme for $P$, and here the constraint function is not necessarily bilinear.

It can be seen that the moment-based approach appears to display better tractability properties. In many cases, the resulting DRCCP problem can be formulated as a tractable second-order conic or semidefinite programming. However, the moment-based approach is relied on the assumption that some certain conditions on the moments are known exactly but that the additional information about the true distribution is ignored. Further, the
moment-based ambiguity set in general cannot converge to the true distribution even more empirical data points are available to estimate moments, which is due to the fact that the moment conditions are insufficient to describe the entire shape of the true distribution. Thus, the resulting worst-case distribution sometimes yields over-conservative decisions. An attractive alternative is to construct the ambiguity set as a ball in the space of probability distributions by using a probability distance function such as the Wasserstein metric. Based on [5,6], we should be aware that the Wasserstein ambiguity sets offer powerful out-of-sample performance guarantees and asymptotic consistency, and also enable the decision makers to control conservativeness of the distribution uncertainty by tuning the radius. In addition, the problem of determining the worst-case expectation over a Wasserstein ambiguity set can in fact be computed efficiently via convex optimization techniques for numerous loss functions of practical interest.

Recently, there are many successful developments on DRCCP problems with Wasserstein ambiguity sets [3,10–13,26,27]. Most existing results on the tractability of DRCCP problems with Wasserstein ambiguity sets are restricted to the case of affine uncertain constraints. For instance, the authors in [3] derived exact mixed-integer conic reformulations for single DRCCP as well as joint DRCCP with right-hand side uncertainty. In [27], the author showed that joint DRCCP is mixed-integer representable by introducing big-M parameters and additional binary variables. He also derived tractable outer and inner approximations of the set $Z_D$. The authors in [13] provided exact reformulations for single DRCCP under discrete support and derived deterministic approximations for single DRCCP under continuous support. To the best of our knowledge, the case of a broader family of uncertain constraints is largely untouched. This kind of optimization problem is of great interest, because the uncertainty might not be inherently affine in many applications. One exception that we are aware of is [12], where the authors studied CVaR approximation of single DRCCP with a general class of constraint functions. They considered many constraint functions that are convex in the decision variable and then replace the set $Z_D$ with convex CVaR approximation. They also presented tractable reformulation of the CVaR approximation when the constraint function is the maximum of functions that are affine in both the decision variable and the uncertainty, and the support of the uncertainty is a polyhedron. Moreover, when the constraint function is concave in the uncertainty, they showed that a central cutting surface algorithm [15,17] for semi-infinite programmings can be used to compute an approximately optimal solution of the CVaR approximation of single DRCCP. Differently from elaborating the tractable approximation base on the special structure of DRCCP in [12], the authors in [26] proposed a DC approximation method for general DRCCP without some special structures and gave the convergence analysis of it under certain weaker conditions. Subsequently, they applied this method to a special case via a sequence of approximated Wasserstein ambiguity sets $\{P_N\}$ centred around the empirical distribution, which can be solved by a recent DC algorithm, and here the constraint function is not necessarily bilinear.

1.3. Contributions, structure, and notations

In this paper, we consider joint DRCCP problems with a general class of convex uncertain constraints under Wasserstein ambiguity sets. Specifically, our main contributions are summarized as below:
(1) We develop the worst-case CVaR approximation of the set $Z_D$ for more complicated and general joint DRCCP, which turns out to be essentially exact under certain conditions. It is proved that the resulting approximation is not convex in general since it involves biconvex constraints. We then derive a convex relaxation of the proposed biconvex approximation by constructing new decision variables which allows us to eliminate biconvex terms.

(2) We transform the convex relaxation of the proposed biconvex approximation into tractable conic programming reformulation when the constraint function is quadratic convex in the uncertainty and concave in the decision variable, and the support of the uncertainty is polyhedron or ellipsoid.

(3) We demonstrate that for joint binary DRCCP, the proposed biconvex approximation admits a tractable mixed-integer convex reformulation when the constraint function is affine in both the decision variable and the uncertainty.

The remainder of the paper is organized as follows. Section 2 presents an exact reformulation of the set $Z_D$. Section 3 develops a biconvex approximation of the set $Z_D$ based on the worst-case CVaR constraints. Section 4 reports the numerical results to illustrate the performance of the proposed models. Finally, Section 5 summarizes this paper.

Notation. The following notation is used throughout the paper. We use bold-letters (e.g., $x, A$) to denote vectors or matrices, and use corresponding non-bold letters to denote their components. Random variables are represented by symbols with tildes, while their realizations are denoted by the same symbols without tildes. We let $e_n$ be the all-one vectors with dimension $n$. Given a positive integer $n$, we let $[n] := \{1, 2, \ldots, n\}$, and $\mathbb{R}_{++}^n := \{x \in \mathbb{R}^n : x_i \geq 0, \forall i \in [n]\}$ and $\mathbb{R}_{++}^n := \{x \in \mathbb{R}^n : x_i > 0, \forall i \in [n]\}$. We denote by $\delta_\xi$ the Dirac distribution concentrating unit mass at $\xi \in \mathbb{R}^n$. We denote $(t)_+ = \max\{t, 0\}$ for any given real number $t$. The product of two probability distributions $P_1$ and $P_2$ on $\Xi_1$ and $\Xi_2$, respectively, is the distribution $P_1 \otimes P_2$ on $\Xi_1 \times \Xi_2$. We define the indicator function as $I_A(\xi) = 1$, if $\xi \in A; = 0$, otherwise. Similarly, the characteristic function is defined as $\chi_A(\xi) = 0$, if $\xi \in A; = \infty$, otherwise. Given a norm $\| \cdot \|$ on $\mathbb{R}^n$, the dual norm $\| \cdot \|_*$ is defined by $\| z \|_* := \sup_{\| \xi \| \leq 1} z^\top \xi$. The space of symmetric matrices of dimension $n$ is denoted by $\mathbb{S}^n$. For any two matrices $X, Y \in \mathbb{S}^n$, the relation $X \succeq Y(X \succ Y)$ implies that $X - Y$ is positive semidefinite (positive definite). The space of positive semidefinite (or positive definite) matrices of dimension $n$ is denoted by $\mathbb{S}_+^n$ (or $\mathbb{S}_++^n$). The inner product of two matrices $X, Y \in \mathbb{R}^{m \times n}$ is denoted by $(X, Y) = \text{tr}(XY) = \sum_{i=1}^m \sum_{j=1}^n x_{ij}y_{ij}$. Additional notations will be introduced as needed.

2. Exact reformulation

In this section, we develop a deterministic reformulation of the set $Z_D$. The derivation of the exact reformulation utilizes strong duality result which is introduced in Lemma 2.1.

We first review the well-known strong duality result from [5], which can be applied to formulate the worst-case chance constraint into its dual form, and indeed by the proof of strong duality theorem in [5], we present another equivalent dual reformulation in Lemma 2.1. To keep this paper self-contained, we present the proof of this lemma here.
Lemma 2.1 (Dual Reformulation, adapted from Theorem 4.2 in [5]): If assumption (A3) holds and \( l(x, \xi) \) is proper, i.e. \( l(x, \xi) < +\infty \) for at least one \( \xi \) and \( l(x, \xi) > -\infty \) for any \( \xi \) in \( \Xi \), then for any given \( \delta > 0 \), the worst-case expectation \( L(x) = \sup_{P \in P_W} \mathbb{E}_P[l(x, \tilde{\xi})] \) equals the optimal value of the following convex program

\[
\inf_{\lambda, s} \left\{ \lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i \right\} \tag{6a}
\]

\[
s.t. \quad \sup_{\xi \in \Xi} \left( l(x, \xi) - \lambda \|\xi - \xi^i\| \right) \leq s_i, \quad \forall i \in [N], \tag{6b}
\]

\[
\lambda \geq 0, \tag{6c}
\]

where \( \lambda \) and \( s \) are decision variables, furthermore, \( \lambda \geq 0 \) is the dual variable for Wasserstein metric constraint \( \int_{\Xi \times \Xi} \|\xi - \xi^i\| \Pi(d\xi, d\xi) \leq \delta \).

**Proof:** By the definition of the Wasserstein metric, the worst-case expectation \( L(x) \) can be rewritten as

\[
\sup_{P, \Pi} \int_{\Xi} l(x, \xi) P(d\xi) \tag{7a}
\]

\[
s.t. \quad P \in \mathcal{M}(\Xi), \tag{7b}
\]

\[
\Pi \in \mathcal{M}(\Xi \times \Xi), \tag{7c}
\]

\[
\int_{\Xi} \|\xi - \xi^i\| \Pi(d\xi, d\xi) \leq \delta, \tag{7d}
\]

where \( \Pi \) is a joint distributionally of \( \tilde{\xi} \) and \( \tilde{\xi}^i \) with marginals \( P \) and \( \mathbb{P}_{\tilde{\xi}} \) respectively.

Note that according to the law of total probability, we can decompose the transportation plan \( \Pi \) for moving the probability mass from \( P \) to \( \mathbb{P}_{\tilde{\xi}} \) as \( \Pi = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi^i} \otimes P_i \), which follows from the fact that any joint probability distribution \( \Pi \) of \( \tilde{\xi} \) and \( \tilde{\xi}^i \) can be constructed from the marginal distribution \( \mathbb{P}_{\tilde{\xi}} \) of \( \tilde{\xi} \) and the conditional distributions \( P_i \) of \( \tilde{\xi} \) given \( \tilde{\xi} = \xi^i, i = 1, 2, \ldots, N \). Thus, problem (7a) equals the optimal value of the generalized moment problem as below,

\[
\sup_{P_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi} l(x, \xi) P_i(d\xi) \tag{8a}
\]

\[
s.t. \quad \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi} \|\xi - \xi^i\| P_i(d\xi) \leq \delta. \tag{8b}
\]

Subsequently, we use Lagrangian duality result to reformulate its dual problem as

\[
L(x) = \sup_{P_i \in \mathcal{M}(\Xi)} \inf_{\lambda \geq 0} \left\{ \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi} l(x, \xi) P_i(d\xi) - \lambda \left( \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi} \|\xi - \xi^i\| P_i(d\xi) - \delta \right) \right\} \tag{9a}
\]
\[
\inf_{\lambda \geq 0} \sup_{\mathcal{P}_i \in \mathcal{M}(\Xi)} \left\{ \lambda \delta + \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi} \left( l(x, \xi) - \lambda \| \xi - \xi^i \| \right) \mathcal{P}_i(d\xi) \right\}
\]

(9b)

\[
= \inf_{\lambda \geq 0} \left\{ \lambda \delta + \frac{1}{N} \sum_{i=1}^{N} \sup_{\xi \in \Xi} \left( l(x, \xi) - \lambda \| \xi - \xi^i \| \right) \right\},
\]

(9c)

where the third equality follows from the fact that \( \mathcal{M}(\Xi) \) contains all the Dirac distributions \( \delta_{\xi^i}, i = 1, 2, \ldots, N \) supported on \( \Xi \).

Here, we show that the second equality in (9a) is valid. Since \( l(x, \xi) \) is proper, then the equivalence of (9a) and (9c) or (9b) is proved case by case.

Assume first that \( l(x, \xi) \) is finite for any \( \xi \in \Xi \). Then, by exploiting a generalization of a well-known strong duality result for moment problems [22, Proposition 3.4], the equivalence holds for any \( \delta > 0 \). Assume next that there exists \( \xi' \in \Xi \) such that \( l(x, \xi') = \infty \). Then, (9c) is infeasible as all of the inner subproblems \( \sup_{\xi \in \Xi} \) are unbounded. In addition, we note that in (8a), \( \int_{\Xi} \| \xi - \xi^i \| \delta_{\xi^i}(d\xi) = 0 < \delta, i = 1, 2, \ldots, N \), and hence the mixture distribution \( \mathcal{P}'_i = (1 - t)\delta_{\xi^i} + t\delta_{\xi'} \) is a feasible solution to (8a) for some \( t \in (0, 1] \) due to the continuity of the norm metric \( \| \cdot \| \). This allows us to conclude that

\[
\sup_{\mathcal{P}_i \in \mathcal{M}(\Xi)} \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi} l(x, \xi) \mathcal{P}_i(d\xi) \geq \frac{1}{N} \sum_{i=1}^{N} \int_{\Xi} l(x, \xi) \mathcal{P}'_i(d\xi) = \infty,
\]

which implies that the prime problem (8a) is unbounded.

Thus, the optimal value of the dual problem (9c) or (9b) coincides with \( L(x) \), that is, the second equality in (9a) turns out to be true.

By introducing auxiliary variables \( s_i, i = 1, 2, \ldots, N \), (9c) can be expressed as

\[
2 \inf_{\lambda, s} \lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i
\]

(10a)

\[
\text{s.t. } \sup_{\xi \in \Xi} \left( l(x, \xi) - \lambda \| \xi - \xi^i \| \right) \leq s_i, \quad \forall i \in [N],
\]

(10b)

\[
\lambda \geq 0,
\]

(10c)

and thus the claim follows.

Next, we can represent the indicator function as a pointwise maximum of a finite number of concave functions by Lemma 2.2 due to [5,13]. This observation is useful for establishing the main result of this section.

**Lemma 2.2:** The indicator function \( \mathbb{I}_{\{l(x, \xi) < 0, \exists t \in [T]\}}(\xi) \) can be rewritten as the pointwise maximum of a finite number of concave functions, which is defined as

\[
\mathbb{I}_{\{l(x, \xi) < 0, \exists t \in [T]\}}(\xi) = \max \left\{ 1 - \chi_{\{l(x, \xi) < 0\}}(\xi), \ldots, 1 - \chi_{\{l(x, \xi) < 0\}}(\xi), 0 \right\},
\]

(11a)

where for any \( t \in [T] \)
Note that for any \( \chi_{\{ f_t(x, \xi) < 0 \}} (\xi) = \begin{cases} 0, & \text{if } f_t(x, \xi) < 0 \\ \infty, & \text{otherwise} \end{cases} \) (11b)

which is the characteristic function of the open convex set defined by \( f_t(x, \xi) < 0 \).

We develop an equivalent reformulation of the set \( Z_D \) in the next theorem which applies to joint DRCCP with more general convex uncertain constraints.

**Theorem 2.1:** The feasible set \( Z_D \) is equivalent to

\[
Z_D = \begin{cases} \lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i \leq \epsilon, \\
G_t^f(z_{it}, \eta_{it}, x) + 1 - z_{it}^T \xi^i - s_i \leq 0, \forall i \in [N], \forall t \in T(x), \\
\|z_{it}\|_{\ast} - \lambda \leq 0, \forall i \in [N], \forall t \in T(x), \\
\lambda \geq 0, \eta_{it} \geq 0, s_i \geq 0, \forall i \in [N], \forall t \in T(x), \end{cases} \tag{12a}
\]

where \( G_t^f(z_{it}, \eta_{it}, x) = \sup_{\xi \in \Xi} [z_{it}^T \xi - \eta_{it} f_t(x, \xi)] \) and \( T(x) = \{ t \in [T] : \exists \xi \in \Xi, f_t(x, \xi) < 0 \} \).

**Proof:** Note that

\[
Z_D := \left\{ x \in \mathbb{R}^n : \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[ \mathbb{I}_{\{ f_t(x, \xi) < 0, \exists t \in [T] \}} \left( \xi \right) \right] \leq \epsilon \right\}.
\]

Therefore, by Lemma 2.1, the left-hand side of the constraint defining \( Z_D \) can be rewritten as

\[
\inf_{\lambda, s} \lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i \tag{13a}
\]

subject to

\[
\sup_{\xi \in \Xi} \left( \mathbb{I}_{\{ f_t(x, \xi) < 0, \exists t \in [T] \}} (\xi) - \lambda \| \xi - \xi^i \| \right) \leq s_i, \quad \forall i \in [N], \tag{13b}
\]

\[
\lambda \geq 0. \tag{13c}
\]

Then, by using the decomposability of \( \mathbb{I}_{\{ f_t(x, \xi) < 0, \exists t \in [T] \}} (\xi) \) into its constituents as denoted in (11a), we obtain that problem (13a) can be reformulated as

\[
\inf_{\lambda, s} \lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i \tag{14a}
\]

subject to

\[
\sup_{\xi \in \Xi} \left( 1 - \chi_{\{ f_t(x, \xi) < 0 \}} - \lambda \| \xi - \xi^i \| \right) \leq s_i, \quad \forall i \in [N], \forall t \in [T] \tag{14b}
\]

\[
\sup_{\xi \in \Xi} \left[ 0 - \lambda \| \xi - \xi^i \| \right] \leq s_i, \quad \forall i \in [N] \tag{14c}
\]

\[
\lambda \geq 0. \tag{14d}
\]

Note that for any \( i \in [N] \) and \( t \in [T] \), we have

\[
\sup_{\xi \in \Xi} \left( 1 - \chi_{\{ f_t(x, \xi) < 0 \}} - \lambda \| \xi - \xi^i \| \right) = \sup_{\xi \in \Xi} \left( 1 - \chi_{\{ f_t(x, \xi) < 0 \}} - \sup_{\| z_{it} \|_{\ast} \leq \lambda} \left( z_{it}^T (\xi - \xi^i) \right) \right)
\]
\[= \sup_{\xi \in \mathbb{S}} \inf_{\|z_{it}\|_* \leq \lambda} \left( 1 - \chi_{\{f_t(x, \xi) < 0\}} - z_{it}^\top (\xi - \xi^i) \right)\]

\[= \inf_{\|z_{it}\|_* \leq \lambda} \left( z_{it}^\top \xi^i + \sup_{\xi \in \mathbb{S}} \left( 1 - \chi_{\{f_t(x, \xi) < 0\}} - z_{it}^\top (\xi) \right) \right),\]

where the first equality is due to the definition of the dual norm and the third equality follows from the general minimax theorem [24, Corollary 3.3], which can be applied because the objective function is convex (linear) in \(z_{it}\) and concave in \(\xi\) and both sets, \(\mathbb{S}\) and \(\{z_{it} \in \mathbb{R}^m : \|z_{it}\|_* \leq \lambda\}\), are closed and convex, indeed, the set \(\{z_{it} \in \mathbb{R}^m : \|z_{it}\|_* \leq \lambda\}\) is also compact for any finite \(\lambda \geq 0\).

Additionally, the inequality (14c) implies that \(s_i \geq 0\) since \(\lambda \geq 0\) and \(\xi^i \in \mathbb{S}\) for any \(i \in [N]\).

Subsequently, we perform variable substitution in which we replace \(z_{it}\) with \(-z_{it}\), and thus problem (14a) can be expressed as

\[
\min_{\lambda, s, z} \lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i
\]

s.t. \(\sup_{\xi \in \mathbb{S}} \left[ z_{it}^\top \xi + 1 - \chi_{\{f_t(x, \xi) < 0\}} (\xi) \right] - z_{it}^\top \xi^i \leq s_i, \forall i \in [N], \forall t \in [T],\) (15b)

\(\|z_{it}\|_* \leq \lambda, \quad \forall i \in [N], \quad \forall t \in [T],\) (15c)

\(\lambda \geq 0, s_i \geq 0, \quad \forall i \in [N].\) (15d)

Furthermore, the optimization problem \(\sup_{\xi \in \mathbb{S}} [z_{it}^\top \xi + 1 - \chi_{\{f_t(x, \xi) < 0\}} (\xi)]\) in (15b) can be rewritten as

\[
\sup_{\xi \in \mathbb{S}} z_{it}^\top \xi + 1
\]

s.t. \(f_t(x, \xi) < 0\) (16b)

for any \(i \in [N]\) and \(t \in T(x)\).

Hence, for any \(i \in [N]\) and \(t \in T(x)\), we use Lagrangian duality result to reformulate the problem (16a) as

\[
\sup_{\xi \in \mathbb{S}} \sup_{\eta_{it} \geq 0} z_{it}^\top \xi + 1 - \eta_{it} f_t(x, \xi)
\]

\(= \inf_{\eta_{it} \geq 0} \sup_{\xi \in \mathbb{S}} \left[ z_{it}^\top \xi + 1 - \eta_{it} f_t(x, \xi) \right],\) (17b)

where the first equality follows from the fact that for any given \(x, f_t(x, \xi)\) and the objective function \(z_{it}^\top \xi + 1\) are both continuous in \(\xi\), \(\mathbb{S}\) is a nonempty closed set, so that we can replace “\(<\)” by “\(\leq\)” without effect on the supremum. Besides, the last equality is because there exists \(\xi \in \mathbb{S}\) such that \(f_t(x, \xi) < 0\) since \(t \in T(x)\), which implies Slater’s condition is satisfied, and hence the strong duality holds.
Thus, for any $i \in [N]$ and $t \in T(\mathbf{x})$, (15b) can be rewritten as
\[
\exists \eta_{it} \geq 0, \sup_{\xi \in \Xi} \left[ z_{it}^\top \xi + 1 - \eta_{it} f_t (\mathbf{x}, \xi) \right] - z_{it}^\top \xi i \leq s_i,
\]
which is equivalent to (12b).

**Remark 2.1:** We note that since $f_t (\mathbf{x}, \xi)$ is convex in $\xi$, then each function $G_{ft} (z_{it}, \eta_{it}, \mathbf{x})$ in constraints (12b) is equivalent to maximizing a concave objective function on a closed convex set $\Xi$.

**Remark 2.2:** The exact reformulation (12) of the set $Z_D$ is not convex because the index set $T(\mathbf{x})$ depends on $\mathbf{x}$ and each function $G_{ft} (z_{it}, \eta_{it}, \mathbf{x})$ in constraints (12b) is not convex in general. Therefore, we attempt to investigate the tractability of the set $Z_D$ by establishing conditions under which $T(\mathbf{x})$ can be replaced by $[T]$ and $G_{ft} (z_{it}, \eta_{it}, \mathbf{x})$ can be convex for any $i \in [N]$ and $t \in [T]$.

## 3. The worst-case CVaR approximation

In this section, we investigate a well-known CVaR approach developed by [18], which can provide a convex inner approximation for the robust single chance constraint (1c) when $f(\mathbf{x}, \xi)$ is concave in $\mathbf{x}$. To be more specific, in [18], the authors constructed a special class of convex inner approximations depending on a class of generating functions $\psi(z)$. By choosing the best generating function to be the piecewise linear function defined as $\psi^*(z) = (1 + z)_+$, the corresponding tightest approximation is established, which is in fact equivalent to the CVaR framework.

We first recall the definitions of VaR and CVaR due to [20]. For a given measurable function $L (\hat{\xi}) : \mathbb{R}^m \to \mathbb{R}$, let $\mathbb{P}$ be its probability distribution and the risk level $\epsilon \in (0, 1)$, then the CVaR at level $\epsilon$ with respect to $\mathbb{P}$ is defined as
\[
\text{CVaR}_{1-\epsilon} (L (\hat{\xi})) := \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_\mathbb{P} \left[ (L (\hat{\xi}) - \beta)_+ \right] \right\},
\]
and the VaR is the $(1 - \epsilon)$-quantile of the distribution of $L (\hat{\xi})$, which is defined as
\[
\text{VaR}_{1-\epsilon} (L (\hat{\xi})) := \inf_{s \in \mathbb{R}} \left\{ s : \mathbb{P} (L (\hat{\xi}) \leq s) \geq 1 - \epsilon \right\}.
\]

It is seen that $\text{VaR}_{1-\epsilon} (L (\hat{\xi}))$ is a minimizer of the right-hand side of the definition of $\text{CVaR}_{1-\epsilon} (L (\hat{\xi}))$, and thus it always holds that $\text{CVaR}_{1-\epsilon} (L (\hat{\xi})) \geq \text{VaR}_{1-\epsilon} (L (\hat{\xi}))$. As this implication holds for any probability distribution and loss function, we conclude that
\[
\text{CVaR}_{1-\epsilon} (-f (\mathbf{x}, \hat{\xi})) \leq 0 \implies \mathbb{P} (-f (\mathbf{x}, \hat{\xi}) \leq 0) \geq 1 - \epsilon,
\]
which further implies that
\[
\sup_{\mathbb{P} \in \mathcal{P}} \text{CVaR}_{1-\epsilon} (-f (\mathbf{x}, \hat{\xi})) \leq 0 \implies \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} (-f (\mathbf{x}, \hat{\xi}) \leq 0) \geq 1 - \epsilon.
\]
Therefore, the worst-case CVaR constraint constitutes a convex inner approximation for the distributionally robust single chance constraint (1c) when $f(\mathbf{x}, \xi)$ is concave in $\mathbf{x}$. 
The above discussion motivates us to reformulate the distributionally robust joint chance constraint \((1c)\) as
\[
Z_D = \left\{ x \in \mathbb{R}^n : \inf_{\mathbb{P} \in \mathcal{P}} \left\{ \tilde{\xi} : \max_{t \in [T]} \left\{ \alpha_t \left( -f_t \left( x, \tilde{\xi} \right) \right) \right\} \leq 0 \right\} \geq 1 - \epsilon \right\}
\] (18)
for any \(\alpha \in \mathbb{R}_+^T\). Note that (18) represents a distributionally robust single chance constraint, which can be conservatively approximated by a worst-case CVaR constraint. Therefore, for any \(\alpha \in \mathbb{R}_+^T\), we have
\[
Z_D \supseteq Z_C (\alpha) = \left\{ x \in \mathbb{R}^n : \alpha \in \mathbb{R}_+^T, \sup_{\mathbb{P} \in \mathcal{P}} \left\{ \inf_{\beta \in \mathbb{R}} \left\{ \beta \right. \right. \right. \\
+ \frac{1}{\epsilon} \mathbb{E}_\mathbb{P} \left[ \left( \max_{t \in [T]} \left\{ \alpha_t \left( -f_t \left( x, \tilde{\xi} \right) \right) \right\} - \beta \right) \right. \right. \right. \\
\left. \left. \left. + \right) \right. \right. \right] \right\} \leq 0 \right\}.
\]
Furthermore, we denote \(Z_C\) as
\[
Z_C = \bigcup_{\alpha \in \mathbb{R}_+^T} Z_C (\alpha) = \left\{ x \in \mathbb{R}^n : \alpha \in \mathbb{R}_+^T, \sup_{\mathbb{P} \in \mathcal{P}} \left\{ \inf_{\beta \in \mathbb{R}} \left\{ \beta \right. \right. \right. \\
+ \frac{1}{\epsilon} \mathbb{E}_\mathbb{P} \left[ \left( \max_{t \in [T]} \left\{ \alpha_t \left( -f_t \left( x, \tilde{\xi} \right) \right) \right\} - \beta \right) \right. \right. \right. \\
\left. \left. \left. + \right) \right. \right. \right] \right\} \leq 0 \right\}.
\]
It should be noted that \(Z_C = \bigcup_{\alpha \in \mathbb{R}_+^T} Z_C (\alpha) \subseteq Z_D\), and then in the subsequent conclusion, we show that \(Z_C\) can be reformulated as a disjunction of two sets \(Z_{C_1}\) and \(Z_{C_2}\).

We first review the stochastic min-max theorem due to [23] before presenting the main results.

**Lemma 3.1 (Stochastic Min-max Equality, Theorem 2.1 in [23]):** Let \(\mathcal{A}\) be a nonempty (not necessarily convex) set of probability measures on measurable space \((\Xi, \mathcal{B} (\Xi))\) where \(\Xi \subseteq \mathbb{R}^m\) and \(\mathcal{B} (\Xi)\) is the Borel \(\sigma\)-algebra. Assume that \(\mathcal{A}\) is weakly compact. Let \(T \subseteq \mathbb{R}^n\) be a closed convex set. Consider a function \(\phi : \mathbb{R}^n \times \Xi \rightarrow \mathbb{R}\). Assume that there exists a convex neighbourhood \(V\) of \(T\) such that for any \(t \in V\), the function \(\phi (t, \cdot)\) is measurable, integrable with respect to all \(\mathbb{P} \in \mathcal{A}\), and \(\sup_{\mathbb{P} \in \mathcal{A}} \mathbb{E}_\mathbb{P} [\phi (t, \tilde{\xi})] < \infty\). Further assume that \(\phi (\cdot, \xi)\) is convex on \(V\) for any \(\xi \in \Xi\). Let \(\bar{t} \in \arg \min_{t \in T} \sup_{\mathbb{P} \in \mathcal{A}} \mathbb{E}_\mathbb{P} [\phi (t, \tilde{\xi})]\). Assume that for every \(t\) in a neighbourhood of \(\bar{t}\), the function \(\phi (t, \cdot)\) is bounded and upper-semicontinuous on \(\Xi\) and the function \(\phi (\bar{t}, \cdot)\) is bounded and continuous on \(\Xi\). Then,
\[
\inf_{t \in T} \sup_{\mathbb{P} \in \mathcal{A}} \mathbb{E}_\mathbb{P} \left[ \phi (t, \tilde{\xi}) \right] = \sup_{\mathbb{P} \in \mathcal{A}} \inf_{t \in T} \mathbb{E}_\mathbb{P} [\phi (t, \tilde{\xi})].
\]

We observe that Lemma 3.1 requires the ambiguity set to be weakly compact. This is indeed the case for Wasserstein ambiguity sets constructed from data due to [19].

**Lemma 3.2 (Proposition 3 in [19]):** The Wasserstein ambiguity set \(\mathcal{P}_W\) defined in (5) is weakly compact.
Proposition 3.1: The set  \( Z_C = Z_{C_1} \cup Z_{C_2} \), where

\[
Z_{C_1} = \left\{ x \in \mathbb{R}^n : f_t(x, \xi) \geq 0, \forall \xi \in \Xi, \forall t \in [T] \right\},
\]

and

\[
Z_{C_2} = \left\{ x \in \mathbb{R}^n : \begin{array}{l}
\lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i \leq \epsilon, \\
\|z_{it}^{-1} \xi - \alpha_t(x)\|_n \leq 0, \forall i \in [N], \forall t \in [T], \\
\lambda \geq 0, \alpha_t \geq 0, s_i \geq 0, \forall i \in [N], \forall t \in [T], \\
\end{array} \right\}
\]

where  \( G_{f_t}(z_{it}, \alpha_t, x) = \sup_{\xi \in \Xi} [z_{it}^\top \xi - \alpha_t f_t(x, \xi)] \).

Proof: We separate the proof into three parts.

(i) Note that

\[
Z_C = \left\{ x \in \mathbb{R}^n : \alpha \in \mathbb{R}^T_{++}, \sup_{\beta \in \mathbb{R}} \{ \inf_{\beta \in \mathbb{R}} \left\{ \begin{array}{l}
\beta \\
d + \frac{1}{\epsilon} \mathbb{E}_P \left[ \max_{t \in [T]} \{ \alpha_t \left( -f_t(x, \tilde{\xi}) \right) \} - \beta \right] \right\} \leq 0 \right\} \right\}
\]

\[
= \left\{ x \in \mathbb{R}^n : \alpha \in \mathbb{R}^T_{++}, \inf_{\beta \in \mathbb{R}} \left\{ \beta \\
d + \frac{1}{\epsilon} \mathbb{E}_P \left[ \max_{t \in [T]} \{ \alpha_t \left( -f_t(x, \tilde{\xi}) \right) \} - \beta \right] \right\} \leq 0 \right\}
\]

\[
= \left\{ x \in \mathbb{R}^n : \alpha \in \mathbb{R}^T_{++}, \beta \in \mathbb{R}, \beta \geq 0 \right\}
\]

\[
+ \frac{1}{\epsilon} \sup_{\beta \in \mathbb{R}} \mathbb{E}_P \left[ \max_{t \in [T]} \{ \alpha_t \left( -f_t(x, \tilde{\xi}) \right) \} - \beta \right] \leq \beta \right\}
\]

where the second equality is due to Lemmas 3.1 and 3.2 and the third equality follows from replacing infimum operator with its equivalent ‘existence’ argument.

Then, we prove that \( \beta \leq 0 \). Suppose that \( \beta > 0 \). Since \( \max_{t \in [T]} \{ \alpha_t(-f_t(x, \xi))\} - \beta \geq 0 \) for any \( \xi \in \Xi \), we must have \( \mathbb{E}_P \left[ \max_{t \in [T]} \{ \alpha_t(-f_t(x, \tilde{\xi}))\} - \beta \right] \geq 0 \). Thus, the left-hand side of (23) is strictly positive, which yields a contradiction.

(ii) Now, we show that \( Z_C \subseteq Z_{C_1} \cup Z_{C_2} \). For any \( x \in Z_C \), there exists \( (\alpha, \beta) \in \mathbb{R}^T_{++} \times \mathbb{R}_- \) such that

\[
\beta \mathbb{E} + \sup_{\beta \in \mathbb{R}} \mathbb{E}_P \left[ \max_{t \in [T]} \{ \alpha_t \left( -f_t(x, \tilde{\xi}) \right) \} - \beta \right] \leq 0.
\]

Then, we distinguish whether \( \beta = 0 \) or \( \beta < 0 \).

Case 1. Note that if \( \beta = 0 \), then inequality (24) implies

\[
\sup_{\beta \in \mathbb{R}} \mathbb{E}_P \left[ \max_{t \in [T]} \{ \alpha_t \left( -f_t(x, \tilde{\xi}) \right) \} \right] \leq 0,
\]

Case 2.
which is equivalent to
\[
\inf_{\mathcal{P} \in \mathcal{P}} \mathbb{P} \left[ f_t(x, \xi) \geq 0, \forall t \in [T] \right] = 1 > 1 - \epsilon,
\]
and hence by continuity of each function \( f_t(x, \xi) \), we have \( f_t(x, \xi) \geq 0 \) for any \( \xi \in \mathcal{X} \). Thus, \( x \in Z_{C_1} \).

**Case 2.** On the other hand, if \( \beta < 0 \), then divide (24) by \(-\beta\) and add \( \epsilon \) on both sides, we have
\[
-\frac{1}{\beta} \sup_{\mathcal{P} \in \mathcal{P}} \mathbb{E}_{\mathcal{P}} \left[ \left( \max_{t \in [T]} \left\{ \alpha_t \left( -f_t(x, \xi) \right) \right\} - \beta \right)_+ \right] \leq \epsilon. \tag{26}
\]
Since \( \beta < 0 \), we can redefine \( \alpha_t \) as \( \alpha_t / (-\beta) \) for any \( t \in [T] \). Then, inequality (26) can be rewritten as
\[
\sup_{\mathcal{P} \in \mathcal{P}} \mathbb{E}_{\mathcal{P}} \left[ \left( \max_{t \in [T]} \left\{ \alpha_t \left( -f_t(x, \xi) \right) \right\} + 1 \right)_+ \right] \leq \epsilon. \tag{27}
\]
Hence, by Lemma 2.1, for any given \( \alpha \in \mathbb{R}_{++}^T \), the supremum in the left-hand side of (27) can be reformulated as
\[
\inf_{\lambda, \delta, \mathcal{S}} \lambda \delta + \frac{1}{N} \sum_{i=1}^N s_i \tag{28a}
\]
subject to
\[
\sup_{\xi \in \mathcal{X}} \left( \left( \max_{t \in [T]} \left\{ \alpha_t \left( -f_t(x, \xi) \right) \right\} + 1 \right)_+ - \lambda \|\xi - \zeta\| \right) \leq s_i, \quad \forall i \in [N], \tag{28b}
\]
\[
\lambda \geq 0, \tag{28c}
\]
Since
\[
\left( \max_{t \in [T]} \left\{ \alpha_t \left( -f_t(x, \xi) \right) \right\} + 1 \right)_+ = \max \left\{ \max_{t \in [T]} \left\{ \alpha_t \left( -f_t(x, \xi) \right) \right\} + 1, 0 \right\} = \max \{ \alpha_1 \left( -f_1(x, \xi) \right) + 1, \ldots, \alpha_T \left( -f_T(x, \xi) \right) + 1, 0 \},
\]
then by using similar arguments as in the proof of Theorem 2.1, for any given \( \alpha \in \mathbb{R}_{++}^T \), problem (28a) is equivalent to
\[
\inf_{\lambda, \delta, \mathcal{S}, \mathcal{Z}} \lambda \delta + \frac{1}{N} \sum_{i=1}^N s_i \tag{29a}
\]
subject to
\[
\sup_{\xi \in \mathcal{X}} \left[ z_{it}^T \xi - \alpha_{it} f_t(x, \xi) + 1 \right] - z_{it}^T \zeta_i \leq s_i, \quad \forall i \in [N], \quad \forall t \in [T], \tag{29b}
\]
\[
\|z_{it}\|_{\ast} \leq \lambda, \quad \forall i \in [N], \quad \forall t \in [T], \tag{29c}
\]
\[
\lambda \geq 0, \quad s_i \geq 0, \quad \forall i \in [N], \tag{29d}
\]
which implies that \( x \in Z_{C_2} \).

(iii) Now, we show that \( Z_{C_1} \cup Z_{C_2} \subseteq Z_C \). Similarly, given \( x \in Z_{C_1} \cup Z_{C_2} \). If \( x \in Z_{C_1} \), then we let \( \beta = 0 \), \( \alpha = e \), thus \( x \in Z_C \). If \( x \in Z_{C_2} \), there exists \( (\lambda', s', \zeta', \alpha', x) \) which
satisfies the constraints in (20). Since $\alpha' > 0$, which could be confirmed in the following Corollary 3.1, and hence let $\beta = -1$, $\alpha = \alpha'$ in (23). Then, by strong duality result introduced in Lemma 2.1, we have $x \in Z_C$. \hfill $\blacksquare$

**Remark 3.1:** To solve the inner approximation of optimization problem (1a) (i.e., $\min_{x \in S \cap Z_C} c^T x$), we can optimize $c^T x$ over $S \cap Z_{C_1}$ and $S \cap Z_{C_2}$ separately, then choose the smallest value.

**Remark 3.2:** We observe that each function $G_{it}(z_{it}, \alpha_t, x)$ is merely biconvex, but not jointly convex in $x$ and $\alpha_t$. Then, the left-hand sides of the constraint system (20) are biconvex in $\alpha$ and $(\lambda, s, z, x)$, i.e. the ya re convex in $(\lambda, s, z, x)$ for any given $\alpha \in \mathbb{R}^T_+$, and also convex in $\alpha$ for any given $(\lambda, s, z, x)$. Thus, optimization problem $\min_{x \in S \cap Z_{C_2}} c^T x$ is non-convex.

Next, we prove that in the constraint system (20), $\alpha_t$ must be strictly positive for any $t \in [T]$.

**Corollary 3.1:** For any $x$ satisfying (20), we must have $\alpha_t > 0$ for any $t \in [T]$.

**Proof:** Suppose that we let $\alpha_{t_0} = 0$ for some $t_0 \in [T]$, then from (20b), we have

$$G_{i_{t_0}}(z_{i_{t_0}}, 0, x) + 1 - z_{i_{t_0}}^T \xi^i = \sup_{\xi \in \Xi} z_{i_{t_0}}^T \xi + 1 - z_{i_{t_0}}^T \xi^i = \sup_{\xi \in \Xi} \left[ z_{i_{t_0}}^T (\xi - \xi^i) \right] + 1 \leq s_i, \quad \forall i \in [N].$$

We note that in (20c), $\|z_{i_{t_0}}\|_* \leq \lambda$ for any $i \in [N]$, we may thus conclude that

$$\inf_{\|z_{i_{t_0}}\|_* \leq \lambda} \left[ z_{i_{t_0}}^T (\xi - \xi^i) \right] + 1 = - \sup_{\|z_{i_{t_0}}\|_* \leq \lambda} \left[ -z_{i_{t_0}}^T (\xi - \xi^i) \right] + 1 \leq s_i, \quad \forall \xi \in \Xi, \forall i \in [N],$$

which is equivalent to

$$-\lambda \|\xi^i - \xi\| + 1 \leq s_i, \quad \forall \xi \in \Xi, \forall i \in [N].$$

Therefore, according to (20a), we have

$$\lambda \delta - \frac{1}{N} \sum_{i=1}^{N} \lambda \|\xi^i - \xi\| + 1 \leq \lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i \leq \epsilon, \quad \forall \xi \in \Xi,$$

which can be rewritten as

$$\lambda \delta - \frac{1}{N} \sum_{i=1}^{N} \lambda \inf_{\xi \in \Xi} \|\xi^i - \xi\| + 1 \leq \lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i \leq \epsilon. \quad (30)$$

Note that $\inf_{\xi \in \Xi} \|\xi^i - \xi\| = 0$ for any $i \in [N]$, thus inequalities (30) yields a contradiction to the fact that $\epsilon < 1$. \hfill $\blacksquare$
Meanwhile, we also show that \( \alpha_t \) could be bounded for any \( t \in [T] \).

**Corollary 3.2:** If \( S \) is compact and \( [T] = T(x) \) for any \( x \in Z_{C_2} \), then there exists an \( M \in \mathbb{R}^T_+ \) such that \( \alpha_t \leq M_t \) for any \( t \in [T] \).

**Proof:** We note that the statement that \( [T] = T(x) \) for any \( x \in Z_{C_2} \) implies given \( x \in S \cap Z_{C_2} \), for any \( t \in [T] \), there exists \( \xi \in \Xi \) such that \( f_t(x, \xi) < 0 \).

Since constraints (20b) can be rewritten as

\[
\sup_{\xi \in \Xi} \left[ z_i^T \xi - \alpha_{\xi} f_t(x, \xi) \right] \leq - \left( 1 - z_i^T \xi^i - s_i \right), \quad \forall i \in [N], \quad \forall t \in [T],
\]

furthermore, which is equivalent to

\[
z_i^T \xi - \alpha_{\xi} f_t(x, \xi) \leq - \left( 1 - z_i^T \xi^i - s_i \right), \quad \forall \xi \in \Xi, \quad \forall i \in [N], \quad \forall t \in [T].
\]

Since \( S \) is compact and \( \Xi \) is closed, we have \( f_t(x, \xi) \) must be finite. Therefore, when \( f_t(x, \xi) < 0 \), we obtain

\[
\alpha_t \leq \frac{1}{f_t(x, \xi)} \left[ z_i^T \xi + 1 - z_i^T \xi^i - s_i \right], \quad \forall i \in [N], \quad \forall t \in [T].
\]

Thus, one can find an upper bound \( M \in \mathbb{R}^T_+ \) such that \( \alpha_t \leq M_t \) for any \( t \in [T] \). \( \blacksquare \)

We can easily observe that \( Z_{C_2} \) is quite similar to the constraint system (12) except that \( G_{f_t}(z_i, \alpha_t, x) \) in (20b) is different from \( G_{f_t}(z_i, \eta_i, x) \) in (12b), and the index set \( T(x) \) is equal to \( [T] \) in (12). In what follows, we show that \( Z_C \) is equivalent to \( Z_D \) under certain conditions.

**Theorem 3.1:** (i) Let \( C := \{ x \in Z_D : \exists t \in [T], \inf_{\xi \in \Xi} f_t(x, \xi) \geq 0 \} \) and assume that \( C = \emptyset \). (ii) Assume that \( \epsilon \in (0, \frac{1}{N}] \). (iii) For any given \( x \in Z_D \), there exists \((\lambda', s', z', \eta', x)\) which satisfies the constraints in (12) and assumes that \( G_{f_t}(z_i', \eta_i', x) = z_i^T \xi^i - \eta_i f_t(x, \xi^i) \). Then \( Z_{C_2} = Z_C = Z_D \).

**Proof:** We note that \( Z_{C_1} \cup Z_{C_2} = Z_C \subseteq Z_D \). Thus, we only need to prove that \( Z_D \subseteq Z_C \). Since \( C = \emptyset \), then \( [T] = T(x) = \{ t \in [T] : \exists \xi \in \Xi, f_t(x, \xi) < 0 \} \) for any \( x \in Z_D \), and further \( Z_{C_1} = \emptyset \). Hence, the index set \( T(x) \) is equal to \( [T] \) in (12), which implies that \( Z_D \) can be rewritten as

\[
Z_D = \left\{ x \in \mathbb{R}^p : \begin{align*}
\lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i &\leq \epsilon, \\
G_{f_t}(z_i, \eta_i, x) + 1 - z_i^T \xi^i - s_i &\leq 0, \forall i \in [N], \forall t \in [T], \\
\|z_i\|_{\infty} - \lambda &\leq 0, \forall i \in [N], \forall t \in [T], \\
\lambda &\geq 0, \forall t \in [N], \forall t \in [T].
\end{align*} \right\}
\]

Indeed, for any given \( x \in Z_D \), there exists \((\lambda', s', z', \eta', x)\) which satisfies the constraints in (31). We first show that \( \lambda' > 0 \). Suppose that \( \lambda' = 0 \), then for any \( i \in [N] \) and \( t \in [T] \), we obtain \( z_i = 0 \). Thus, by (31b), we have

\[
\sup_{\xi \in \Xi} \left[ -\eta_i f_t(x, \xi) \right] + 1 - s_i \leq 0,
\]

which implies that for any \( i \in [N] \), \( s_i \geq 1 \) since \( \sup_{\xi \in \Xi} [ -\eta_i f_t(x, \xi) ] \geq 0 \) from the fact that \( C = \emptyset \). Consequently, by (31a), we have \( \epsilon \geq 1 \), a contradiction.
We next show that $s_i' < 1$ if $N \epsilon \leq 1$. Suppose that we let $s_{i_0}' \geq 1$ for some $i_0 \in [N]$. Then from (31a), we have

$$\lambda' \delta \leq \epsilon - \frac{1}{N} \sum_{i=1}^{N} s_i' = -\frac{1}{N} \sum_{i \in [N] \setminus \{i_0\}} s_i' + \frac{1}{N} (N \epsilon - s_{i_0}') \leq 0,$$

where the second inequality is due to $N \epsilon \leq 1$ and $s_{i_0}' \geq 1$. This yields a contradiction to the fact that $\delta > 0$ and $\lambda' > 0$. Therefore, in (31b), we must have $G_f(\zeta'_{it}, \eta_{it}, x) - \zeta'^{\top}_{it} \xi_i < 0$.

Note that $G_f(\zeta'_{it}, \eta_{it}, x) - \zeta'^{\top}_{it} \xi_i = \zeta'^{\top}_{it} \xi_i - \eta_{it} f_t(x, \xi_i) = -\eta_{it} f_t(x, \xi_i) < 0$, which implies that $\eta_{it}$ can be chosen as a sufficiently large positive number $M$ at optimality. Hence, $(\lambda', s', \zeta', \eta', x)$ satisfies the constraints in (20), i.e. $x \in Z_C = Z_{C_2}$, where $\eta_{it}' = \alpha_{it}' = M$ for any $i \in [N]$ and $t \in [T]$.

**Remark 3.3:** Note that in Theorem 3.1, if $f_t(x, \xi)$ is bilinear and $\Xi = \mathbb{R}^m$, then the set $Z_D$ is equivalent to its CVaR approximation when the risk parameter $\epsilon$ is small enough.

The following theorem is derived by constructing two new decision variables which allows us to eliminate biconvex terms and proving that the new formulation is a convex relaxation of the set $Z_{C_2}$.

**Theorem 3.2:** The set $Z_{C_2}$ can be outer approximated by

$$Z_{C_2} = \left\{ x \in \mathbb{R}^n : \begin{align*}
\|v_{it}\|_\delta + \frac{1}{N} \sum_{i=1}^{N} q_{it} &\leq \epsilon \alpha_t, \forall i \in [N], \forall t \in [T], \\
G_f(v_{it}, 1, x) + \alpha_t - v_{it}^{\top} \xi_i - q_{it} &\leq 0, \forall i \in [N], \forall t \in [T], \\
\alpha_t > 0, q_{it} &\geq 0, \forall i \in [N], \forall t \in [T],
\end{align*} \right\}$$

which is a convex set.

**Proof:** By Proposition 3.1 and Corollary 3.1, the set $Z_{C_2}$ can be expressed as

$$Z_{C_2} = \left\{ x \in \mathbb{R}^n : \begin{align*}
\lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i &\leq \epsilon, \\
\alpha_t \sup_{\xi \in \Xi} \left[ \frac{\| \xi \|_\delta}{\alpha_t} - f_t(x, \xi) \right] + 1 - \alpha_t (\frac{\| \xi \|_\delta}{\alpha_t}) &\geq 0, \forall i \in [N], \forall t \in [T], \\
\|z_{it}\|_\delta - \lambda &\leq 0, \forall i \in [N], \forall t \in [T], \\
\lambda &\geq 0, \alpha_t > 0, s_i &\geq 0, \forall i \in [N], \forall t \in [T].
\end{align*} \right\}$$

Note that for any $i \in [N]$ and $t \in [T]$, $\|z_{it}\|_\delta \leq \lambda$ is equivalent to $\lambda \geq \max_{i \in [N], t \in [T]} \|z_{it}\|_\delta$. This fact along with the first inequality constraint $\lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i \leq \epsilon$ implies that it is optimal to set $\lambda = \max_{i \in [N], t \in [T]} \|z_{it}\|_\delta$ for any $x \in Z_{C_2}$, which finally allows us to reformulate $Z_{C_2}$ as

$$Z_{C_2} = \left\{ x \in \mathbb{R}^n : \begin{align*}
\sup_{\xi \in \Xi} \left[ \frac{\| \xi \|_\delta}{\alpha_t} - f_t(x, \xi) \right] + 1 - \alpha_t (\frac{\| \xi \|_\delta}{\alpha_t}) &\geq 0, \forall i \in [N], \forall t \in [T], \\
\|z_{it}\|_\delta &\leq \epsilon, \forall i \in [N], \forall t \in [T], \\
- \frac{s_i}{\alpha_t} &\leq 0, \forall i \in [N], \forall t \in [T], \\
\alpha_t > 0, s_i &\geq 0, \forall i \in [N], \forall t \in [T].
\end{align*} \right\}$$

Subsequently, we introduce two new decision variables $v_{it} = \frac{z_{it}}{\alpha_t}$ and $q_{it} = \frac{s_i}{\alpha_t}$. But we cannot recover the original $s_i$ by using $q_{it}$, and thus the resulting set turns out to be a relaxation
of the set $Z_{C_2}$. Then, the set $Z_{C_2}$ can be outer approximated by

$$
\tilde{Z}_{C_2} = \left\{ \mathbf{x} \in \mathbb{R}^n : \begin{array}{l}
\|v_{lt}\|_\delta + \frac{1}{N} \sum_{i=1}^{N} q_{lt} \leq \frac{\alpha_t}{\alpha_t}, \forall i \in [N], \forall t \in [T], \\
\sup_{\xi \in \Xi} \left[ v_{lt}^T \xi - f_i (x, \xi) \right] + \frac{1}{\alpha_t} - v_{lt}^T \xi^i \\
- q_{lt} \leq 0, \forall i \in [N], \forall t \in [T], \\
\alpha_t > 0, q_{lt} \geq 0, \forall i \in [N], \forall t \in [T].
\end{array} \right\}
$$

We now perform variable substitution in which we replace $\frac{1}{\alpha_t}$ by $\alpha_t$, which yields the following reformulation of $\tilde{Z}_{C_2}$

$$
\tilde{Z}_{C_2} = \left\{ \mathbf{x} \in \mathbb{R}^n : \begin{array}{l}
\|v_{lt}\|_\delta + \frac{1}{N} \sum_{i=1}^{N} q_{lt} \leq \varepsilon \alpha_t, \forall i \in [N], \forall t \in [T], \\
\sup_{\xi \in \Xi} \left[ v_{lt}^T \xi - f_i (x, \xi) \right] + \alpha_t - v_{lt}^T \xi^i \\
- q_{lt} \leq 0, \forall i \in [N], \forall t \in [T], \\
\alpha_t > 0, q_{lt} \geq 0, \forall i \in [N], \forall t \in [T].
\end{array} \right\}
$$

and thus the claim follows.

\[\blacksquare\]

**Remark 3.4:** It is easy to check that if there is a single uncertain constraint (i.e., $T = 1$), then $\tilde{Z}_{C_2} = Z_{C_2}$.

We now present tractable conic reformulations of the set $\tilde{Z}_{C_2}$ when the constraint function is quadratic convex in the uncertainty and concave in the decision variable, and the support of the uncertainty is polyhedron or ellipsoid. In addition, we provide the detailed proofs in Appendix 1.

**Proposition 3.2:** If the support set $\Xi$ is polyhedral, i.e. $\Xi = \{ \xi \in \mathbb{R}^m : a_k^T \xi \leq d_k, a_k \in \mathbb{R}^m, d_k > 0, k = 1, 2, \ldots, l \}$, and $f_i(x, \xi) = \xi^T x + (A^i, \xi \xi^T) + (b^i)^T x + h^i$ for any $t \in [T]$, where $A^i \succeq 0$. Then

$$
\tilde{Z}_{C_2} = \left\{ \mathbf{x} \in \mathbb{R}^n : \begin{array}{l}
\|v_{lt}\|_\delta + \frac{1}{N} \sum_{i=1}^{N} q_{lt} \leq \varepsilon \alpha_t, \forall i \in [N], \forall t \in [T], \\
u_{lt} - [(b^i)^T x + h^i] + \alpha_t - v_{lt}^T \xi^i \leq q_{lt}, \forall i \in [N], \forall t \in [T], \\
U(u_{lt}, v_{lt}, x, v^i) \succeq 0, \forall i \in [N], \forall t \in [T], \\
\alpha_t > 0, v^i_k \geq 0, q_{lt} \geq 0, u_{lt} \in \mathbb{R}, \forall i \in [N], \forall t \in [T], \\
\forall k \in [l],
\end{array} \right\}
$$

where

$$
U(u_{lt}, v_{lt}, x, v^i) = \begin{bmatrix}
A^i \\
-\frac{1}{2} (v^i_{lt} - x - \sum_{k=1}^{l} v^i_k a_k)^\top \\
-\frac{1}{2} (v^i_{lt} - x - \sum_{k=1}^{l} v^i_k a_k)
\end{bmatrix}.
$$

**Proposition 3.3:** If the support set $\Xi$ is ellipsoidal, i.e. $\Xi = \{ \xi \in \mathbb{R}^m : (\xi - \xi_0)^\top W^{-1} (\xi - \xi_0) \leq 1 \}$, where $W > 0$, and $f_i(x, \xi) = \xi^T x + (A^i, \xi \xi^T) + (b^i)^T x + h^i$ for any
Proposition 3.4: If the support set \( \Xi \) is ellipsoidal, i.e. \( \Xi = \{ \xi \in \mathbb{R}^m : (\xi - \xi_0)^\top W^{-1}(\xi - \xi_0) \leq 1 \} \), where \( W > 0 \), and \( f_t(\xi, \mathbf{x}) = w_t(\xi)^\top \mathbf{x} \) for any \( t \in [T] \), where each component \( w_{ij}(\xi) \) of \( w_t(\xi) \) is quadratic in \( \xi \), i.e. it has the form \( w_{ij}(\xi) = \xi_i^\top W_{ij}\xi_i + r_{ij}\xi_i + h_{ij} \), where \( W_{ij} \geq 0 \). Then

\[
\hat{Z}_{C_2} = \begin{cases} 
\mathbf{x} \in \mathbb{R}^n : & \|\mathbf{v}_t\|_* \delta + \frac{1}{N} \sum_{i=1}^N q_{it} \leq \epsilon \alpha_t, \forall i \in [N], \forall t \in [T], \\
u_{it} - (b^\top_i + h^\top_i) + \alpha_{t} - v_{it}^\top \xi_i^t \leq q_{it}, \forall i \in [N], \forall t \in [T], \\
\mathbf{u}(\mathbf{v}_t, \mathbf{v}_{it}, \mathbf{x}, v_{it}) \geq 0, \forall i \in [N], \forall t \in [T], \\
\alpha_t > 0, v_{it} \geq 0, q_{it} \geq 0, u_{it} \in \mathbb{R}, \forall i \in [N], \forall t \in [T], 
\end{cases}
\]

where

\[
\mathbf{u}(\mathbf{v}_t, \mathbf{v}_{it}, \mathbf{x}, v_{it}) = \left[ -\frac{1}{2} (2v_{it} W^{-1}_i + v_{it} - x^\top) \right].
\]

Remark 3.5: We observe that by Propositions 3.2–3.4, \( \hat{Z}_{C_2} \) has a manifestly tractable representation in terms of Linear Matrix Inequalities (LMIs), which can be solved directly by using the powerful convex optimization solvers.

We prove that for joint binary DRCCP, i.e. \( S \subseteq \{0, 1\}^n \), the set \( S \cap Z_{C_2} \) can be expressed as a mixed-integer convex reformulation when the constraint function is affine in both the decision variable and the uncertainty.

Theorem 3.3: Suppose that \( S \subseteq \{0, 1\}^n \), \( f_t(\mathbf{x}, \xi) = (A^t\mathbf{x} + a^t)^\top \xi + (b^t)^\top \mathbf{x} + h^t \) for any \( t \in [T] \), and \( \alpha \) in (20) can be upper bounded by a vector \( \mathbf{M} \) for any \( \mathbf{x} \in S \). Consider a convex set

\[
\hat{Z}_{C_2} = \begin{cases} 
\mathbf{x} \in \mathbb{R}^n : & \frac{\lambda}{N} \sum_{i=1}^N s_i \leq 0,
\end{cases}
\]

where \( \hat{f}_t((\mathbf{z}_t, y^t), \xi) = (A^t y^t + a^t \alpha_t)^\top \xi + (b^t)^\top y^t + h^t \alpha_t \) for any \( t \in [T] \), then

\[
S \cap \hat{Z}_{C_2} = S \cap Z_{C_2} \subseteq S \cap Z_D.
\]
**Proof:** If each function $f_i(x, \xi) = (A^i x + a^i)^T \xi + (b^i)^T x + h^i$, then for any $t \in [T]$ and $i \in [N]$, we have

$$G_{f_t}(z_{ht}, \alpha_t, x) = \sup_{\xi \in \Xi} \left\{ z_{ht}^T \xi - \alpha_t \left[ (A^i x + a^i)^T \xi + (b^i)^T x + h^i \right] \right\}.$$ 

Now, we define new variables $y^t$ as $y^t = \alpha_t x$ for any $t \in [T]$. Since $\alpha_t \leq M_t$ for any $t \in [T]$, and hence by McCormick inequalities due to [16], we obtain

$$0 \leq y^t_r \leq M_t x_r, \quad \alpha_t - M_t (1 - x_r) \leq y^t_r \leq \alpha_t,$$

which is exact for any $x \subseteq [0, 1]^n$. Thus, we have $S \cap \hat{Z}_{C_2} = S \cap Z_{C_2}$. 

**Remark 3.6:** We note that by Theorem 3.3, to optimize over $S \cap Z_{C_2}$, we only need to optimize over $S \cap \hat{Z}_{C_2}$, which is a mixed-integer convex set.

**Remark 3.7:** In Theorem 3.3, we have assumed that one can find an upper bound $M$, and indeed Corollary 3.2 also provides a sufficient condition for the existence of the vector $M$.

We show how to find the proper upper bound $M$ on the variable $\alpha$ for any $x \in S \cap Z_{C_2}$ in the following example.

**Example 3.1:** Let $\Xi = \mathbb{R}^m$ in Theorem 3.3, then the set $Z_{C_2}$ can be rewritten as

$$Z_{C_2} = \left\{ x \in \mathbb{R}^n : \begin{array}{l}
\lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i \leq \epsilon, \\
1 - s_i \leq \alpha_t \left( (A^i x + a^i)^T \xi^i + (b^i)^T x + h^i \right), \\
\forall i \in [N], \forall t \in [T] , \\
\alpha_t \|A^i x + a^i\|_\infty - \lambda \leq 0, \forall t \in [T], \\
\lambda \geq 0, \alpha_t \geq 0, s_i \geq 0, \forall i \in [N], \forall t \in [T].
\end{array} \right\} \quad (37a)$$

For any given $t \in [T]$, we add up all of constraints (37b) and obtain the inequality as below:

$$N - \sum_{i=1}^{N} s_i \leq \alpha_t \left( (A^i x + a^i)^T \sum_{i=1}^{N} \xi^i + N(b^i)^T x + Nh^i \right), \quad (38)$$

Furthermore, by constraint (37a), we note that

$$N + N(\lambda \delta - \epsilon) \leq \alpha_t \left( (A^i x + a^i)^T \sum_{i=1}^{N} \xi^i + N(b^i)^T x + Nh^i \right), \quad (39)$$

which implies that

$$N + N(\lambda \delta - \epsilon) \leq \alpha_t \left( \|A^i\|_1 \sum_{i=1}^{N} \xi^i + (a^i)^T \sum_{i=1}^{N} \xi^i + N \|b^i\|_1 + Nh^i \right). \quad (40)$$

We observe that if $\mu \leq 0$, then inequality (40) can be rewritten as $N + N(\lambda \delta - \epsilon) \leq 0$, which yields a contradiction to the fact that $\lambda \delta \geq 0$, and hence we must have $\mu > 0$, where $\mu = \|A^i\|_1 \sum_{i=1}^{N} \xi^i + (a^i)^T \sum_{i=1}^{N} \xi^i + N \|b^i\|_1 + Nh^i$. 


Since for any \( i \in [N] \), \( s_i \geq 0 \), then by constraints (37a) and (37c), we have \( \alpha_t \leq \frac{\gamma}{\delta} \leq \frac{\epsilon}{\delta} \), where \( \gamma = \min_{x \in \{0,1\}^n} \|A'x + a'\|_\infty \).

Thus, we obtain \( \frac{N + N(\lambda \delta - \epsilon)}{\mu} \leq \alpha_t \leq \frac{\epsilon}{\delta} \), and then for any \( t \in [T] \), we let \( M_t = \frac{\epsilon}{\delta} \).

4. Numerical results

In this section, we present two groups of numerical studies to demonstrate the computational effectiveness of the proposed formulations, one is to compare the out-of-sample performance of the approximation approach proposed in Theorem 3.2 with that of the sample average approximation (SAA) method in the context of a transportation decision problem, the other is to compare the approximation model proposed in Theorem 3.3 with exact Big-M model proposed in [27] so as to evaluate the optimality gap of the approximation model relative to the true optimality in the context of a binary knapsack problem.

4.1. Distributionally robust chance-constrained transportation decision problem

We now consider a transportation decision problem, which is a classical problem that has been extensively studied in many literatures. The following example is an adaptation from [30]. Nonetheless, for completeness, we provide the following description of the problem. Given a set of facilities indexed by \( K = \{k : k = 1, \ldots, m \} \) and a set of customer locations indexed by \( J = \{j : j = 1, \ldots, n \} \). The total production quantity at each facility \( k \) is \( a_k \). Further assume that each facility has a normalized production capacity of \( L_{\text{high}} \), i.e. \( a_k \leq L_{\text{high}} \). Similarly, the total demand quantity from each customer location \( j \) is \( b_j \). Assume that the total production quantity equals to the total demand quantity, and they must be larger than a minimum demand quantity \( L_{\text{low}} \), i.e. \( \sum_{k=1}^{m} a_k = \sum_{j=1}^{n} b_j \geq L_{\text{low}} \). We aim to determine how to transport these products after all the ordered products are manufactured by the facilities in order to minimize the total cost. To formulate the problem, we define the uncertain parameter \( \tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_m)^T \) to be the vector of the unit transportation cost, where \( \tilde{\xi}_k = (\tilde{\xi}_{k1}, \tilde{\xi}_{k2}, \ldots, \tilde{\xi}_{kn}) \), \( k = 1, 2, \ldots, m \), and \( x_{kj} \) to be the volume of transport from point \( k \) to \( j \):

\[
\begin{align*}
    f(\xi, a, b) = \min_{x_{kj}} & \sum_{k=1}^{m} \sum_{j=1}^{n} \tilde{\xi}_{kj} x_{kj} \\
    \text{s.t.} & \sum_{j=1}^{n} x_{kj} = a_k, \quad \forall k \in K, \\
    & \sum_{k=1}^{m} x_{kj} = b_j, \quad \forall j \in J, \\
    & x_{kj} \geq 0, \quad \forall k \in K, \quad \forall j \in J.
\end{align*}
\]

It should be noted that the unit transportation cost \( \tilde{\xi} \) supported on a rectangle of the form \( \Xi = \{\xi \in \mathbb{R}^{nm} : 0 \leq \xi \leq d\} \) are only revealed at the second stage, then to ensure
that the total transportation cost is low with high probability, the first-stage decision can be formulated as a DRCCP problem:

\[
\begin{align*}
\min_{a, b, z} & \quad z \\
\text{s.t.} & \quad \inf_{P \in \mathcal{P}} \mathbb{P}\left\{ \tilde{\xi} : f(\tilde{\xi}, a, b) \leq z \right\} \geq 1 - \epsilon, \\
& \quad \sum_{k=1}^{m} a_k = \sum_{j=1}^{n} b_j \geq L_{\text{low}}, \\
& \quad 0 \leq a_k \leq L_{\text{high}}, b_j \geq 0, \quad \forall k \in K, \quad \forall j \in J.
\end{align*}
\]

Note that \( z - f(\xi, a, b) \) is convex in the uncertainty \( \xi \) and concave in the decision variable \((a, b, z)\), which obviously satisfies assumption (A1).

Next, we review different approaches to construct \( \mathcal{P} \) from \( N \) sample data points \( \{\xi^i\}_{i \in [N]} \subseteq \Xi \) generated from the true distribution \( \mathbb{P}_{\text{true}} \). Thus, for the proposed model (41a), we compare the out-of-sample performance of the distributionally robust approach based on \( 1 \)-Wasserstein ball (denoted as DRW Model) with that of the classical sample average approximation (SAA) method (denoted as SAA Model).

For DRW Model, we use the Wasserstein ambiguity set under assumption (A3) with \( 1 \)-norm as distance metric. Then, problem (41a) can therefore be amenable to the worst-case CVaR approximation method discussed in Section 3. We thus obtain the approximation model by Theorem 3.2 as follows:

\[
\begin{align*}
\min_{a, b, x, y, v, q, \alpha, z} & \quad z \\
\text{s.t.} & \quad \delta |v_{ir}| + \frac{1}{N} \sum_{i=1}^{N} q_i \leq \epsilon \alpha, \quad \forall i \in [N], \quad \forall r \in [mn], \\
& \quad (y^i)^\top d - z + \alpha - v_{i}^\top \xi^i - q_i \leq 0, \quad \forall i \in [N], \\
& \quad v_i + x^i \leq y^i, \quad \forall i \in [N], \\
& \quad \sum_{j=1}^{n} x_{kj}^i = a_k, \quad \sum_{k=1}^{m} x_{kj}^i = b_j, \quad \forall k \in [m], \quad \forall j \in [n], \\
& \quad \forall i \in [N], \\
& \quad \sum_{k=1}^{m} a_k = \sum_{j=1}^{n} b_j \geq L_{\text{low}}, \\
& \quad 0 \leq a_k \leq L_{\text{high}}, b_j \geq 0, \quad x_{kj}^i \geq 0, \quad y^i \geq 0, \\
& \quad q_i \geq 0, \alpha > 0, \quad \forall k \in [m], \quad \forall j \in [n], \quad \forall i \in [N].
\end{align*}
\]

For SAA Model, we set \( \mathcal{P} = \{ \mathbb{P}_{\xi} \} \), which corresponds to a Wasserstein ball centred at the empirical distribution with the Wasserstein radius \( \epsilon = 0 \). Then, problem (41a) simply reduces to the corresponding SAA problem, which can be expressed as the following
linear optimization problem

\[
\min_{a,b,z} \quad z
\]

\[
\text{s.t.} \quad \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{\{z - f(\xi^i, a, b) \geq 0\}} \geq 1 - \epsilon, \quad (43a)
\]

\[
\sum_{k=1}^{m} a_k = \sum_{j=1}^{n} b_j \geq L_{low}, \quad (43b)
\]

\[
0 \leq a_k \leq L_{high}, \quad b_j \geq 0, \quad \forall k \in [m], \quad \forall j \in [n], \quad (43c)
\]

where \(\{\xi^i\}_{i \in [N]} \subseteq \Xi\) are sample data points generated from the true distribution \(P_{\text{true}}\). Furthermore, according to [21], we can convert this into a large mixed-integer linear programming (MILP) as follows:

\[
\min_{a,b,x^i,s,z} \quad z
\]

\[
\text{s.t.} \quad z - \xi^i \top x^i - M_i(s_i - 1) \geq 0, \quad \forall i \in [N], \quad (44a)
\]

\[
\frac{1}{N} \sum_{i=1}^{N} s_i \geq 1 - \epsilon, \quad (44b)
\]

\[
\sum_{j=1}^{n} x^i_{kj} = a_k, \quad \sum_{k=1}^{m} x^i_{kj} = b_j, \quad \forall k \in [m], \quad \forall j \in [n], \quad \forall i \in [N], \quad (44c)
\]

\[
\sum_{k=1}^{m} a_k = \sum_{j=1}^{n} b_j \geq L_{low}, \quad (44d)
\]

\[
\mathbf{s} \in \{0,1\}^N, \quad 0 \leq a_k \leq L_{high}, \quad b_j \geq 0, \quad x^i_{kj} \geq 0, \quad \forall k \in [m], \quad \forall j \in [n], \quad \forall i \in [N], \quad (44e)
\]

where \(M_i\) is a sufficiently large positive constant.

In the subsequent tests, we set \(m = 4, n = 6, L_{low} = L_{high} = 2\), and \(d = 8e_{mn}\). We assume that the true distribution \(P_{\text{true}}\) of the unit transportation cost \(\tilde{\xi}\) is lognormal, which is designed in the following manner. That is, \(\tilde{\xi}_{kj} = \exp(\tilde{\eta}_{kj})\), where \(\tilde{\eta}_{kj}, k \in [m], j \in [n]\), represent jointly normally distributed random variables with the mean \(\mu\) drawn uniformly from \([0,1]^{mn}\) and the covariance matrix \(\Sigma = \text{Diag}(\sigma) C \text{Diag}(\sigma)\), where \(C \in S_{+}^{mn}\) is a random correlation matrix and \(\sigma = 2e_{mn}\) is the vector of standard deviations. Each problem instance is solved with CPLEX 12.10 using the YALMIP interface on a desktop with a 4.10 GHz processor and 32GB RAM.

To assess the out-of-sample performance of different data-driven methods mentioned above, we conduct out-of-sample experiments to compare the reliability of the performance guarantees for these two models. Then we compute the optimal solution \((a^*_1, b^*_1, z^*_1)\) of DRW Model by solving problem (42a). Similarly, the optimal solution \((a^*_2, b^*_2, z^*_2)\) of SAA Model is obtained by solving problem (44a). Moreover, the out-of-sample performance is
Figure 1. The reliability of DRW model as a function of $\delta$ with different sample size $N = 10$ and $160$.

measured by the following chance constraint

$$\mathbb{P}_{\text{true}} \{ \tilde{\xi} : z^* - f(\tilde{\xi}, a^*, b^*) \geq 0 \} \geq 1 - \epsilon,$$  \hspace{1cm} \text{(45a)}

which can be estimated at high accuracy using 20,000 test samples generated from $\mathbb{P}_{\text{true}}$ by solving another SAA problem

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{I} \{ z^* - f(\xi^i, a^*, b^*) \geq 0 \} \geq 1 - \epsilon.$$  \hspace{1cm} \text{(45b)}

Thus, in the out-of-sample evaluation, for a given optimal solution $(a^*, b^*, z^*)$ of DRW Model or SAA Model, checking whether it is feasible for chance constraint (45a) simplifies to calculating its reliability using (45b) and then comparing with the risk level $1 - \epsilon$. In the following Subsections 4.1.1 and 4.1.2, we study how the Wasserstein radius $\delta$ and the samplesize $N$ affect the reliability of the optimal solutions for DRW Model and SAA Model.

4.1.1. Impact of the Wasserstein radius $\delta$ on reliability

In this subsection, we compute the error bars between the 20% and 80% quantiles as well as the mean values of reliability for DRW Model under different radius $\delta \in \{0.01, 0.04, 0.07, 0.10, 0.13, 0.16, 0.19\}$ with the sample size $N = 10$ and $160$, respectively, averaged across 20 independent random instances. Besides, we set the risk level to $\epsilon = 0.10$.

Figure 1 displays the reliability of the optimal solutions for DRW Model as a function of $\delta$. We observe that the reliability of DRW Model increases as $\delta$ grows. This is because the larger the radius $\delta$ is, the more distributions the Wasserstein ball $\mathcal{P}$ includes, and accordingly a larger uncertainty distribution family can capture the more uncertainty and can be a truer response for the all uncertainties. Moreover, it should be noted that from Figure 1(a), as long as the proper radius is chosen, DRW Model is able to generate a feasible solution even if the size of the sample data set is very small. On the other hand, Figure 1(b) shows that when the sample data size is large enough, DRW Model can obtain a feasible solution even with the small radius.
Figure 2. The reliability of DRW model and SAA model as a function of $N$ with different radius $\delta = 0.10$ and 0.19.

4.1.2. Impact of the sample size $N$ on reliability

In this subsection, we compute the error bars between the 20% and 80% quantiles as well as the mean values of reliability for DRW Model and SAA Model under different sample size $N \in \{10, 20, 40, 80, 160, 320, 640\}$ with the Wasserstein radius $\delta = 0.10$ and 0.19, respectively, averaged across 20 independent random instances. Besides, we set the risk level to $\epsilon = 0.10$.

Figure 2 depicts the reliability of the optimal solutions for DRW Model and SAA Model as a function of $N$. We see that the reliability of both models tends to increase as the sample size becomes larger. Furthermore, Figure 2(b) shows that DRW Model can provide a high-quality reliable solution with a proper choice of radius $\delta$ even when the sample data points are very limited. However, SAA Model yields a poor reliability in situation where $N$ is small. In addition, the error bars visualize that DRW Model is significantly more stable than SAA Model. These results demonstrate that DRW Model is capable of returning reliable and stable solutions.

4.2. Distributionally robust multidimensional knapsack problem

For the evaluation purpose, we study distributionally robust multidimensional knapsack problem (DRMKP) \cite{2,25} with binary decision variables. The following notations are adopted for binary DRMKP. We consider $T$ knapsacks and $n$ items, moreover, $c_i$ represents the value of item $i$ for any $i \in [n]$, $\tilde{\xi}_t = (\tilde{\xi}_{t1}, \ldots, \tilde{\xi}_{tn})^\top$ represents the vector of random item weights supported on $\Xi_t$ in knapsack $t$, and $b_t > 0$ represents the capacity limit of knapsack $t$ for any $t \in [T]$. In addition, the decision variable $x_i \in \{0, 1\}$ represents the proportion of $i$th item to be picked for any $i \in [n]$ and we let $x \in S := \{0, 1\}^n$. Furthermore, we use the Wasserstein ambiguity sets under assumptions (A2) and (A3). With the notations above, binary DRMKP can be formulated as

\begin{align}
\max_x & \quad c^\top x \\
\text{s.t.} & \quad x \in \{0, 1\}^n .
\end{align}
\[
\inf_{\mathcal{P} \in \mathcal{P}} \mathbb{P}\left\{ \xi_t \in \Xi_t : \xi^T_t x \leq b_t, \forall t \in [T] \right\} \geq 1 - \epsilon, \tag{46c}
\]

where constraint (46c) is to guarantee that the worst-case probability that the capacity of each knapsack should be satisfied is at least \(1 - \epsilon\).

The following example demonstrates an application of Theorem 3.3. As observed in [27], the author derived an exact Big-M approach to solve it.

**Example 4.1:** Consider binary DRMKP (46a) with joint chance constraint \((T > 1)\). Suppose the Wasserstein ambiguity set be defined as

\[
\mathcal{P} = \left\{ \mathbb{P} : \mathbb{P}\left\{ \xi \in \Xi \right\} = 1, \inf_{\mathbb{Q}} \left\{ \int_{\Xi \times \Xi} \|\xi - \zeta\|_2 \mathbb{Q}(d\xi, d\zeta) \right\} \leq \delta \right\},
\]

where \(\Xi = \prod_{t \in [T]} \Xi_t\) and \(\Xi_t = \mathbb{R}^n\).

Then, by Theorem 3.3, binary DRMKP (46a) can be approximated by mixed-integer second-order cone programming (MISOCP) as follows:

\[
\begin{align*}
\text{max} & \quad c^T x, \quad \tag{47a} \\
\text{s.t.} & \quad x \in \{0, 1\}^n, \quad \tag{47b} \\
& \quad \lambda \delta + \frac{1}{N} \sum_{i=1}^{N} s_i \leq \epsilon, \quad \tag{47c} \\
& \quad -\alpha_t b_t + (y^t)^T \xi^t_x + 1 - s_i \leq 0, \quad \forall i \in [N], \quad \forall t \in [T], \quad \tag{47d} \\
& \quad 0 \leq y^t_r \leq \frac{\epsilon}{\delta} x_r, \quad \alpha_t - \frac{\epsilon}{\delta} (1 - x_r) \leq y^t_r \leq \alpha_t, \quad \forall r \in [n], \quad \forall t \in [T], \quad \tag{47e} \\
& \quad \|y^t\|_2 - \lambda \leq 0, \quad \forall t \in [T], \quad \tag{47f} \\
& \quad \lambda \geq 0, \alpha_t \geq 0, s_i \geq 0, \quad \forall i \in [N], \quad \forall t \in [T]. \quad \tag{47g}
\end{align*}
\]

It should be noted that according to Example 3.1, \(\alpha_t\) can be upper bounded by \(M_t = \frac{\epsilon}{\delta}\) for all \(t \in [T]\) in problem (47a).

By Example 4.1, we present one group of numerical experiments, i.e. to compare our approximation method proposed in Theorem 3.3 with exact Big-M method proposed in [27].

The problem instances are created as follows. We generate 20 random instances with \(n = 20\) and \(T = 10\). For each instance, we generate \(N \in \{100, 1000\}\) empirical samples \(\{\xi^i\}_{i \in [N]} \subseteq \mathbb{R}^{n \times T}\) from a uniform distribution over a box \([1, 10]^{n \times T}\). For each \(i \in [n]\), we independently generate \(c_i\) from the uniform distribution on the interval \([1, 10]\). In addition, we set \(b_t = 50\) for each \(t \in [T]\). We test these 20 random instances with the risk level \(\epsilon \in \{0.05, 0.10\}\) and the Wasserstein radius \(\delta \in \{0.01, 0.02\}\). All the instances are executed on a desktop with a 4.10 GHz processor and 32GB RAM, while CPLEX 12.10 is used with their default settings. We set the time limit of solving each instance to be 3600 seconds.

The numerical results with sample size \(N = 100\) are displayed in Table 1. We use Big-M Model to denote exact reformulation proposed in [27]. Besides, we use CVaR Model to denote inner approximation proposed in Theorem 3.3. We use ‘Opt.Val’ to denote the
Table 1. Numerical results of Big-M Model proposed in [27] and CVaR Model proposed in Theorem 3.3 for binary DRMKP when sample size $N = 100$.

| $T$ | $n$ | $\epsilon$ | $\delta$ | Instances | Opt. Val | Time | Value | GAP | Time |
|-----|-----|-------------|-----------|-----------|----------|------|-------|-----|------|
| 1   | 20  | 0.05       | 0.01      | 1         | 50.49    | 4.82 | 50.49 | 0.00%| 12.75 |
| 2   | 20  | 0.05       | 0.01      | 2         | 50.32    | 2.32 | 50.32 | 0.00%| 6.29  |
| 3   | 20  | 0.05       | 0.01      | 3         | 49.51    | 7.67 | 49.51 | 0.00%| 8.02  |
| 4   | 20  | 0.05       | 0.01      | 4         | 50.10    | 9.67 | 49.90 | 0.40%| 13.34 |
| 5   | 20  | 0.05       | 0.01      | 5         | 49.55    | 2.18 | 48.25 | 2.63%| 9.52  |
| 6   | 20  | 0.05       | 0.01      | 6         | 44.00    | 3.07 | 43.96 | 0.07%| 7.66  |
| 7   | 20  | 0.05       | 0.01      | 7         | 55.33    | 8.16 | 55.33 | 0.00%| 6.12  |
| 8   | 20  | 0.05       | 0.01      | 8         | 52.64    | 42.29| 52.64 | 0.00%| 9.40  |
| 9   | 20  | 0.05       | 0.01      | 9         | 43.08    | 1.27 | 43.08 | 0.00%| 4.87  |
| 10  | 20  | 0.05       | 0.01      | 10        | 38.81    | 4.32 | 38.81 | 0.00%| 8.68  |
| 11  | 20  | 0.05       | 0.01      | 11        | 54.96    | 3.17 | 54.96 | 0.00%| 8.17  |
| 12  | 20  | 0.05       | 0.01      | 12        | 49.95    | 2.82 | 49.95 | 0.00%| 9.14  |
| 13  | 20  | 0.05       | 0.01      | 13        | 54.41    | 2.57 | 53.86 | 1.00%| 5.34  |
| 14  | 20  | 0.05       | 0.01      | 14        | 47.29    | 2.55 | 47.29 | 0.00%| 5.31  |
| 15  | 20  | 0.05       | 0.01      | 15        | 51.64    | 3.76 | 51.09 | 1.08%| 8.36  |
| 16  | 20  | 0.05       | 0.01      | 16        | 48.41    | 8.46 | 48.41 | 0.00%| 9.49  |
| 17  | 20  | 0.05       | 0.01      | 17        | 52.30    | 37.08| 52.30 | 0.00%| 11.60 |
| 18  | 20  | 0.05       | 0.01      | 18        | 52.61    | 3.54 | 52.61 | 0.00%| 7.66  |
| 19  | 20  | 0.05       | 0.01      | 19        | 52.06    | 26.50| 52.06 | 0.00%| 13.36 |
| 20  | 20  | 0.05       | 0.01      | 20        | 54.49    | 3.86 | 53.93 | 1.02%| 8.19  |
|     |     | Average    |           |           | 50.10    | 9.00 | 49.94 | 0.31%| 8.66  |
| 10  | 20  | 0.05       | 0.02      | 10        | 55.36    | 4.19 | 54.98 | 0.69%| 9.45  |
| 11  | 20  | 0.05       | 0.02      | 11        | 53.90    | 5.66 | 53.90 | 0.00%| 6.15  |
| 12  | 20  | 0.05       | 0.02      | 12        | 49.80    | 2.20 | 49.80 | 0.00%| 7.92  |
| 13  | 20  | 0.05       | 0.02      | 13        | 50.78    | 4.01 | 50.78 | 0.00%| 6.65  |
| 14  | 20  | 0.05       | 0.02      | 14        | 49.16    | 2.95 | 47.68 | 3.01%| 7.81  |
| 15  | 20  | 0.05       | 0.02      | 15        | 49.88    | 3.52 | 49.88 | 0.00%| 7.90  |
| 16  | 20  | 0.05       | 0.02      | 16        | 55.15    | 3.25 | 53.90 | 2.26%| 8.54  |
| 17  | 20  | 0.05       | 0.02      | 17        | 52.17    | 4.07 | 52.17 | 0.00%| 5.27  |
| 18  | 20  | 0.05       | 0.02      | 18        | 49.47    | 2.42 | 47.84 | 3.29%| 5.90  |
| 19  | 20  | 0.05       | 0.02      | 19        | 51.14    | 2.27 | 48.50 | 5.16%| 10.06 |
| 20  | 20  | 0.05       | 0.02      | 20        | 42.80    | 7.40 | 42.80 | 0.00%| 5.28  |
|     |     | Average    |           |           | 50.19    | 5.53 | 49.80 | 0.76%| 7.49  |
|     |     | 1          |           |           | 51.75    | 59.64| 51.75 | 0.00%| 9.41  |
|     |     | 2          |           |           | 44.62    | 199.04| 44.62 | 0.00%| 12.48 |
|     |     | 3          |           |           | 54.03    | 27.20| 54.03 | 0.00%| 10.27 |
|     |     | 4          |           |           | 56.37    | 760.00| 56.37 | 0.00%| 13.75 |
|     |     | 5          |           |           | 50.70    | 69.95| 50.70 | 0.00%| 12.40 |
|     |     | 6          |           |           | 49.40    | 606.17| 49.40 | 0.00%| 10.22 |
|     |     | 7          |           |           | 38.60    | 231.03| 38.60 | 0.00%| 14.38 |
|     |     | 8          |           |           | 52.57    | 457.77| 52.57 | 0.00%| 8.82  |
|     |     | 9          |           |           | 47.88    | 221.65| 47.88 | 0.00%| 8.99  |
| 10  | 20  | 0.10       | 0.01      | 10        | 52.69    | 188.16| 52.69 | 0.00%| 11.76 |
| 11  | 20  | 0.10       | 0.01      | 11        | 50.85    | 91.39| 50.02 | 1.62%| 12.41 |
| 12  | 20  | 0.10       | 0.01      | 12        | 51.91    | 2168.96| 51.83 | 0.16%| 12.71 |

(continued)
Table 1. Continued.

| $T$ | $n$ | $\epsilon$ | $\delta$ | Big-M Model | CVaR Model |
|-----|-----|------------|----------|-------------|------------|
|     |     |            |          | Instances | Opt.Val | Time | Value | GAP | Time |
| 13  | 1  | 47.46      | 75.18    | 47.46      | 0.00%    | 9.66  |
| 14  | 1  | 54.97      | 187.18   | 54.97      | 0.00%    | 8.39  |
| 15  | 1  | 49.93      | 85.41    | 49.93      | 0.00%    | 11.28 |
| 16  | 1  | 53.39      | 395.89   | 53.39      | 0.00%    | 12.74 |
| 17  | 1  | 50.97      | 330.25   | 50.97      | 0.00%    | 14.11 |
| 18  | 1  | 51.03      | 77.50    | 51.03      | 0.00%    | 11.28 |
| 19  | 1  | 46.65      | 239.84   | 46.65      | 0.00%    | 12.15 |
| 20  | 1  | 50.89      | 415.31   | 50.89      | 0.00%    | 11.58 |
|     |    | 50.33      | 344.38   | 50.29      | 0.09%    | 11.44 |
| 1   | 2  | 48.69      | 98.88    | 48.69      | 0.00%    | 6.97  |
| 2   | 2  | 54.14      | 1142.32  | 53.61      | 0.98%    | 13.64 |
| 3   | 2  | 41.95      | 198.40   | 41.95      | 0.00%    | 12.72 |
| 4   | 2  | 53.28      | 19.94    | 53.28      | 0.00%    | 6.32  |
| 5   | 2  | 54.94      | 27.46    | 54.94      | 0.00%    | 7.61  |
| 6   | 2  | 54.52      | 45.11    | 54.52      | 0.00%    | 6.93  |
| 7   | 2  | 55.96      | 739.68   | 55.96      | 0.00%    | 13.04 |
| 8   | 2  | 52.77      | 285.83   | 52.77      | 0.00%    | 13.35 |
| 9   | 2  | 50.60      | 129.35   | 50.60      | 0.00%    | 8.79  |
| 10  | 2  | 47.15      | 138.42   | 47.15      | 0.00%    | 12.25 |
| 11  | 2  | 44.59      | 5.02     | 43.32      | 2.85%    | 7.03  |
| 12  | 2  | 49.85      | 378.80   | 49.62      | 0.46%    | 12.35 |
| 13  | 2  | 55.42      | 68.98    | 55.42      | 0.00%    | 9.79  |
| 14  | 2  | 50.23      | 79.51    | 50.23      | 0.00%    | 12.46 |
| 15  | 2  | 50.47      | 119.03   | 50.06      | 0.80%    | 14.23 |
| 16  | 2  | 47.05      | 271.61   | 47.05      | 0.00%    | 15.00 |
| 17  | 2  | 52.30      | 37.53    | 51.83      | 0.89%    | 8.17  |
| 18  | 2  | 51.85      | 74.08    | 51.53      | 0.60%    | 9.21  |
| 19  | 2  | 54.30      | 265.85   | 54.20      | 0.18%    | 14.08 |
| 20  | 2  | 50.89      | 22.53    | 50.89      | 0.00%    | 8.71  |
|     |    | 51.05      | 207.42   | 50.87      | 0.36%    | 10.63 |

optimal value $v^*$, ‘Value’ to denote the best objective value output from the approximation model (47a), and ‘Time’ to denote the total running time in seconds. Additionally, since we can solve Big-M Model to the optimality, we use ‘GAP’ to denote the optimality gap of the approximation model (47a), which is computed as

$$
\text{GAP} = \frac{|\text{Value} - \text{Opt.Val}|}{\text{Opt.Val}}.
$$

The numerical results with sample size $N = 1000$ are displayed in Table 2. Similarly, we use Big-M Model to denote exact reformulation proposed in [27]. Besides, we use CVaR Model to denote inner approximation proposed in Theorem 3.3. We use ‘LB’ to denote the best lower bounds found by the models that we’re going to test, and ‘Time’ to denote the total running time in seconds. To evaluate the effectiveness of CVaR Model, we use ‘Improvement’ to denote the percentage of differences between the lower bound of the approximation model (47a) and the lower bound of Big-M Model, which is computed as

$$
\text{Improvement} = \frac{\text{LB of Approximation Model} - \text{LB of Exact Model}}{\text{LB of Exact Model}},
$$

where Approximation Model here represents the approximation model (47a). Similarly, Exact Model here represents Big-M Model.
Table 2. Numerical results of Big-M Model proposed in [27] and CVaR Model proposed in Theorem 3.3 for binary DRMKP when sample size $N = 1000$.

| $T$ | $n$ | $\epsilon$ | $\delta$ | Instances | Big-M Model | CVaR Model |
|-----|-----|-------------|-----------|-----------|-------------|------------|
|     |     |             |           |           | LB | Time | LB | Improvement | Time |
| 1   | 40.44 | 3600 | 40.44 | 0.00% | 1171.59 |
| 2   | 40.18 | 3600 | 41.14 | 2.40% | 1032.97 |
| 3   | 43.70 | 3600 | 43.70 | 0.00% | 476.74 |
| 4   | 43.65 | 3600 | 43.65 | 0.00% | 1375.80 |
| 5   | 44.15 | 3600 | 44.15 | 0.00% | 783.29 |
| 6   | 45.48 | 3600 | 45.48 | 0.00% | 637.16 |
| 7   | 47.92 | 3600 | 47.92 | 0.00% | 839.95 |
| 8   | 44.84 | 3600 | 44.84 | 0.00% | 725.06 |
| 9   | 47.00 | 3600 | 47.00 | 0.00% | 555.20 |
| 10  | 45.73 | 3600 | 45.73 | 0.00% | 986.77 |
| 11  | 44.16 | 3600 | 44.16 | 0.00% | 543.22 |
| 12  | 40.11 | 3600 | 40.11 | 0.00% | 990.35 |
| 13  | 46.67 | 3600 | 46.67 | 0.00% | 617.06 |
| 14  | 46.04 | 3600 | 50.57 | 9.83% | 1173.02 |
| 15  | 37.91 | 3600 | 43.65 | 15.12% | 1033.04 |
| 16  | 44.64 | 3600 | 45.13 | 1.10% | 1172.05 |
| 17  | 45.43 | 3600 | 45.43 | 0.00% | 1000.34 |
| 18  | 46.98 | 3600 | 47.15 | 0.36% | 898.53 |
| 19  | 47.78 | 3600 | 48.34 | 1.17% | 979.69 |
| 20  | 39.85 | 3600 | 39.85 | 0.00% | 427.03 |
| Average | 44.13 | 3600 | 44.75 | 1.50% | 870.94 |
| 1   | 39.33 | 3600 | 39.33 | 0.00% | 782.68 |
| 2   | 43.83 | 3600 | 43.83 | 0.00% | 892.32 |
| 3   | 42.46 | 3600 | 42.46 | 0.00% | 1269.51 |
| 4   | 44.00 | 3600 | 44.00 | 0.00% | 780.20 |
| 5   | 45.12 | 3600 | 45.12 | 0.00% | 1458.42 |
| 6   | 43.84 | 3600 | 43.84 | 0.00% | 1028.58 |
| 7   | 45.08 | 3600 | 45.08 | 0.00% | 903.12 |
| 8   | 36.46 | 3600 | 36.46 | 0.00% | 702.42 |
| 9   | 40.09 | 3600 | 40.09 | 0.00% | 947.85 |
| Average | 44.13 | 3600 | 44.75 | 1.50% | 870.94 |
| 11  | 44.16 | 3600 | 44.16 | 0.00% | 543.22 |
| 12  | 40.11 | 3600 | 40.11 | 0.00% | 990.35 |
| 13  | 46.67 | 3600 | 46.67 | 0.00% | 617.06 |
| 14  | 46.04 | 3600 | 50.57 | 9.83% | 1173.02 |
| 15  | 37.91 | 3600 | 43.65 | 15.12% | 1033.04 |
| 16  | 44.64 | 3600 | 45.13 | 1.10% | 1172.05 |
| 17  | 45.43 | 3600 | 45.43 | 0.00% | 1000.34 |
| 18  | 46.98 | 3600 | 47.15 | 0.36% | 898.53 |
| 19  | 47.78 | 3600 | 48.34 | 1.17% | 979.69 |
| 20  | 39.85 | 3600 | 39.85 | 0.00% | 427.03 |
| Average | 42.34 | 3600 | 42.73 | 0.00% | 858.10 |
| 1   | 52.91 | 3600 | 52.91 | 0.00% | 1435.84 |
| 2   | 49.08 | 3600 | 49.21 | 0.27% | 2074.25 |
| 3   | 50.01 | 3600 | 50.81 | 1.60% | 1543.52 |
| 4   | 51.03 | 3600 | 51.03 | 0.00% | 446.97 |
| 5   | 46.46 | 3600 | 46.46 | 0.00% | 1397.00 |
| 6   | 53.46 | 3600 | 53.46 | 0.00% | 1420.57 |
| 7   | 49.66 | 3600 | 49.66 | 0.00% | 1548.56 |
| 8   | 56.37 | 3600 | 56.37 | 0.00% | 1178.06 |
| 9   | 52.02 | 3600 | 52.02 | 0.00% | 722.68 |
| Average | 52.23 | 3600 | 52.23 | 0.00% | 635.57 |
| 10  | 51.39 | 3600 | 51.39 | 0.00% | 813.74 |
| 11  | 51.18 | 3600 | 51.18 | 0.00% | 1020.88 |
| 12  | 50.04 | 3600 | 50.04 | 0.00% | 1286.21 |

(continued)
Table 2. Continued.

| $T$ | $n$ | $\epsilon$ | $\delta$ | Instances | LB | Time | LB | Improvement | Time |
|-----|-----|-------------|----------|-----------|-----|------|----|-------------|------|
| 14  | 47.66 | 3600 | 47.66 | 0.00% | 953.13 |
| 15  | 51.40 | 3600 | 51.40 | 0.00% | 783.22 |
| 16  | 52.45 | 3600 | 52.45 | 0.00% | 595.98 |
| 17  | 51.92 | 3600 | 52.16 | 0.46% | 1296.64 |
| 18  | 53.41 | 3600 | 53.41 | 0.00% | 598.95 |
| 19  | 49.58 | 3600 | 49.58 | 0.00% | 1081.70 |
| 20  | 49.91 | 3600 | 50.59 | 1.36% | 1844.17 |
| Average | 51.11 | 3600 | 51.20 | 0.18% | 1133.88 |
| 1   | 53.85 | 3600 | 53.98 | 0.24% | 1246.58 |
| 2   | 46.97 | 3600 | 47.97 | 2.13% | 867.48 |
| 3   | 48.51 | 3600 | 49.29 | 1.60% | 591.15 |
| 4   | 42.03 | 3600 | 43.71 | 4.00% | 980.55 |
| 5   | 39.83 | 3600 | 41.01 | 2.96% | 1008.77 |
| 6   | 49.22 | 3600 | 49.98 | 1.54% | 746.48 |
| 7   | 47.97 | 3600 | 49.00 | 2.14% | 747.23 |
| 8   | 52.48 | 3600 | 52.48 | 0.00% | 1115.56 |
| 9   | 40.20 | 3600 | 40.20 | 0.00% | 722.10 |
| 10  | 46.24 | 3600 | 46.24 | 0.00% | 882.39 |
| 11  | 50.60 | 3600 | 50.60 | 0.00% | 768.26 |
| 12  | 52.91 | 3600 | 52.91 | 0.00% | 719.34 |
| 13  | 47.99 | 3600 | 48.30 | 0.65% | 649.62 |
| 14  | 52.87 | 3600 | 52.91 | 0.07% | 843.17 |
| 15  | 43.44 | 3600 | 44.16 | 1.65% | 691.86 |
| 16  | 50.68 | 3600 | 51.79 | 2.18% | 710.75 |
| 17  | 45.67 | 3600 | 45.67 | 0.00% | 407.13 |
| 18  | 49.10 | 3600 | 49.10 | 0.00% | 919.86 |
| 19  | 56.23 | 3600 | 57.70 | 2.61% | 763.05 |
| 20  | 51.15 | 3600 | 51.15 | 0.00% | 894.84 |
| Average | 48.40 | 3600 | 48.97 | 1.20% | 813.81 |

From Table 1, we note that CVaR Model can be solved to the optimality within 15 seconds, while Big-M Model often takes longer time to solve. In terms of approximation accuracy, CVaR Model nearly finds the true optimal solutions in most instances. These results demonstrate that CVaR Model is capable of finding near-optimal solutions.

From Table 2, we observe that the total running time of CVaR Model significantly outperforms that of Big-M Model, i.e. the majority of the instances of CVaR Model can be solved within 20 minutes, while Big-M Model cannot be solved within the time limit. In terms of approximation accuracy, we see that CVaR Model can find at least the same feasible solutions as Big-M Model. Additionally, in some instances, CVaR Model can provide slightly better lower bounds than Big-M Model. These results demonstrate the effectiveness of CVaR Model, which scales for large number of sample data points.

As can be seen from Tables 1 and 2, the main reasons for the different numerical performances between Big-M Model and CVaR Model are as follows: (i) CVR Model only involves $O(n)$ binary variables, while Big-M Model involves $O(N + n)$ binary variables. (ii) Big-M Model requires more auxiliary variables than CVaR Model.

5. Conclusion

In this paper, we studied distributionally robust joint chance-constrained programming with convex uncertain constraints under Wasserstein ambiguity set. We proposed an
equivalent reformulation of the set $Z_D$ and then developed the worst-case CVaR approximation of the set $Z_D$ via a system of biconvex constraints, which is naturally hard to solve. Furthermore, a convex relaxation of the proposed approximation can be derived by constructing new decision variables which allows us to eliminate biconvex terms. Once the decision variables are binary and the uncertain constraints are affine, this proposed biconvex approximation is equivalent to a tractable mixed-integer convex programming. Numerical results demonstrated that the proposed models can be solved efficiently. In our study, we assume the uncertain mapping $f_i(x, \xi)$ is convex in $\xi$ and concave in $x$. A future direction is to consider DRCCP problems with a broader family of uncertain mappings, for instance, when each constraint function is quasi-convex in the uncertainty and is concave in the decision variable.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This research was supported by National Natural Science Foundation of China (Grant No. 11271243).

**Notes on contributors**

**Yining Gu,** received the M.S. degree from the School of Mathematics, Tianjin University, Tianjin, China, in 2018. She is currently pursuing the Ph.D. degree with the School of Mathematics, Shanghai University of Finance and Economics, Shanghai, China. Her research interests include stochastic optimization, distributionally robust optimization, and their applications.

**Yanjun Wang,** received the Ph.D. degree from the School of Science, Xi’an Jiaotong University, Xi’an, China, in 2004. She is currently a Professor with the School of Mathematics, Shanghai University of Finance and Economics, Shanghai, China. Her research interests include stochastic optimization, distributionally robust optimization, non-convex optimization, global optimization, and their applications.

**Data availability statement**

Some or all data, models, or code generated or used during the study are available from the first author of the paper by request.

**References**

[1] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, J. Polit. Econ. 81(3) (1973), pp. 637–654.
[2] G.C. Calafiore and L. ElGhaoui, *On distributionally robust chance-constrained linear programs*, J. Optim. Theory Appl. 130(1) (2006), pp. 1–22.
[3] Z. Chen, D. Kuhn, and W. Wiesemann, *Data-driven chance constrained programs over Wasserstein balls*, Oper. Res. (2022). https://doi.org/10.1287/opre.2022.2330.
[4] J. Cheng, E. Delage, and A. Lisser, *Distributionally robust stochastic knapsack problem*, SIAM J. Optim. 24(3) (2014), pp. 1485–1506.
[5] P.M. Esfahani and D. Kuhn, *Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations*, Math. Program. 171(1-2) (2018), pp. 115–166.
[6] R. Gao and A.J. Kleywegt, Distributionally robust stochastic optimization with Wasserstein distance, Math. Oper. Res. (2022). https://doi.org/10.1287/moor.2022.1275.

[7] S. Guo, H. Xu, and L. Zhang, Convergence analysis for mathematical programs with distributionally robust chance constraint, SIAM J. Optim. 27(2) (2017), pp. 784–816.

[8] G.A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann, A distributionally robust perspective on uncertainty quantification and chance constrained programming, Math. Program. 151 (2015), pp. 35–62.

[9] G.A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann, Ambiguous joint chance constraints under mean and dispersion information, Oper. Res. 65(3) (2017), pp. 751–767.

[10] N. Ho-Nguyen, F. Klinc-Karzan, S. Küçükyavuz, and D. Lee, Distributionally robust chance-constrained programs with right-hand side uncertainty under Wasserstein ambiguity, Math. Program. 196 (2022), pp. 641–672.

[11] N. Ho-Nguyen, F. Klinc-Karzan, S. Küçükyavuz, and D. Lee, Strong formulations for distributionally robust chance-constrained programs with left-hand side uncertainty under Wasserstein ambiguity, INFORMS J. Optim. (2022). https://doi.org/10.1287/iijo.2022.0083.

[12] A.R. Hota, A. Cherukuri, and J. Lygeros, Data-driven chance constrained optimization under Wasserstein ambiguity sets, in 2019 American Control Conference (ACC), IEEE, 2019, pp. 1501–1506. https://doi.org/10.23919/ACC.2019.8814677.

[13] R. Ji and M.A. Lejeune, Data-driven distributionally robust chance-constrained optimization with Wasserstein metric, J. Glob. Optim. 79 (2021), pp. 779–811.

[14] B. Kawas and A. Thiele, Short sales in log-robust portfolio management, Eur. J. Oper. Res. 215(3) (2011), pp. 651–661.

[15] F. Luo and S. Mehrotra, Decomposition algorithm for distributionally robust optimization using Wasserstein metric with an application to a class of regression models, Eur. J. Oper. Res. 278(1) (2019), pp. 20–35.

[16] G.P. Mccormick, Computability of global solutions to factorable nonconvex programs: Part I-convex underestimating problems, Math. Program. 10(1) (1976), pp. 147–175.

[17] S. Mehrotra and D. Papp, A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization, SIAM J. Optim. 24(4) (2014), pp. 1670–1697.

[18] A. Nemirovski and A. Shapiro, Convex approximation of chance constrained programs, SIAM J. Optim. 17(4) (2007), pp. 969–996.

[19] A. Pichler and H. Xu, Quantitative stability analysis for minimax distributionally robust risk optimization, Math. Program. 191 (2022), pp. 47–77.

[20] R.T. Rockafellar and S. Uryasev, Optimization of conditional value-at-risk, J. Risk 2 (2000), pp. 21–41.

[21] A. Ruszczynski, Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedra, Math. Program. 93 (2002), pp. 195–215.

[22] A. Shapiro, On duality theory of conic linear problems, in Semi-Infinite Programming, Goberna, M.A. and López, M.A. eds., Springer, Boston, MA, 2001, pp. 135–165.

[23] A. Shapiro and A. Kleywegt, Minimax analysis of stochastic problems, Optim. Methods Softw. 17(3) (2002), pp. 523–542.

[24] M. Sion, On general minimax theorems, Pac. J. Math. 8(1) (1958), pp. 171–176.

[25] Y. Song, J.R. Luedtke, and S. Küçükyavuz, Chance-constrained binary packing problems, INFORMS J. Comput. 26(4) (2014), pp. 735–747.

[26] H. Sun, D. Zhang, and Y. Chen, Convergence analysis and a DC approximation method for Data-driven mathematical programs with distributionally robust chance constraints, 2019. Available at Optimization Online http://www.optimization-online.org/DB_HTML/2019/11/7465.html.

[27] W. Xie, On distributionally robust chance constrained programs with Wasserstein distance, Math. Program. 186 (2021), pp. 115–155.

[28] W. Xie and S. Ahmed, On deterministic reformulations of distributionally robust joint chance constrained optimization problems, SIAM J. Optim. 28(2) (2018), pp. 1151–1182.

[29] W. Xie, S. Ahmed, and R. Jiang, Optimized Bonferroni approximations of distributionally robust joint chance constraints, Math. Program. 191 (2022), pp. 79–112.
which is equivalent to
\[\xi\]

We now discuss the reformulation of the inner subproblem \(\sup_{\xi \in \Xi} \) in (A1). For this problem, there exists \(\xi\) such that \(a_{k}^{\top} \xi < d_{k}, k = 1, 2, \ldots, l\) due to \(d_{k} > 0\), which implies Slater’s condition is satisfied, and hence the strong duality holds.

Thus, we have
\[
\sup_{\xi \in \Xi} \left[ (v_{lt} - x)^{\top} \xi - (A^{\top}, \xi^{\top}) \right] = \inf_{\mu^{\alpha} \geq 0} \sup_{\xi \in \Xi} \left[ (v_{lt} - x)^{\top} \xi - (A^{\top}, \xi^{\top}) - \sum_{k=1}^{l} v_{it}^{a} (a_{k}^{\top} \xi - d_{k}) \right],
\]
which is equivalent to
\[
\min_{\mu^{\alpha} \geq 0, u_{lt} \in \Xi} u_{lt}
\]
\[
\text{s.t. } (v_{lt} - x)^{\top} \xi - (A^{\top}, \xi^{\top}) - \sum_{k=1}^{l} v_{it}^{a} (a_{k}^{\top} \xi - d_{k}) \leq u_{lt}, \quad \forall \xi \in \Xi
\]
which can be further written as
\[
\min_{\mu^{\alpha} \geq 0, u_{lt} \in \Xi} u_{lt}
\]
\[
\text{s.t. }\left[ -\frac{1}{2} (v_{lt} - x - \sum_{k=1}^{l} v_{k} a_{k})^{\top} - \frac{1}{2} (v_{lt} - x - \sum_{k=1}^{l} v_{k} a_{k}) \right] \geq 0
\]
and thus, (A1) is equivalent to
\[
q_{lt} \geq \min_{\mu^{\alpha} \geq 0, u_{lt} \in \Xi} u_{lt} - \left[ (b')^{\top} x + h' \right] + \alpha_{l} - v_{lt}^{\top} \xi^{i}
\]
\[
\text{s.t. }\left[ -\frac{1}{2} (v_{lt} - x - \sum_{k=1}^{l} v_{k} a_{k})^{\top} - \frac{1}{2} (v_{lt} - x - \sum_{k=1}^{l} v_{k} a_{k}) \right] \geq 0.
\]
Then the claim follows.

**Proof of Proposition 3.3:** Note that similar to the proof of Proposition 3.2, (32b) can be rewritten as
\[
q_{lt} \geq \sup_{\xi \in \Xi} \left[ (v_{lt} - x)^{\top} \xi - (A^{\top}, \xi^{\top}) \right] - \left[ (b')^{\top} x + h' \right] + \alpha_{l} - v_{lt}^{\top} \xi^{i}, \quad (A2)
\]
We now discuss the reformulation of the inner subproblem \( \sup_{\xi \in \Xi} \) in (A2). For this problem, there exists \( \xi = \xi_0 \) such that \((\xi - \xi_0)^\top W^{-1}(\xi - \xi_0) < 1\), which implies Slater’s condition is satisfied, and hence the strong duality holds.

Thus, we have
\[
\sup_{\xi \in \Xi} \left( (v_{it} - x)^\top \xi - (A^i, \xi \xi^\top) \right) = \inf_{v_{it} \geq 0} \sup_{\xi \in \Xi} \left[ (v_{it} - x)^\top \xi - (A^i, \xi \xi^\top) - v_{it}((\xi - \xi_0)^\top W^{-1}(\xi - \xi_0) - 1) \right],
\]
which is equivalent to
\[
\min_{v_{it} \geq 0, u_{it} \in \mathbb{R}} u_{it}
\text{ s.t. } (v_{it} - x)^\top \xi - (A^i, \xi \xi^\top) - v_{it}((\xi - \xi_0)^\top W^{-1}(\xi - \xi_0) - 1) \leq u_{it}, \forall \xi \in \mathbb{R}^m,
\]
which can be further written as
\[
\min_{v_{it} \geq 0, u_{it} \in \mathbb{R}} u_{it}
\text{ s.t. } \begin{bmatrix} A^i + v_{it}W^{-1} \\ -\frac{1}{2}(2v_{it}W^{-1}\xi_0 + v_{it} - x)^\top \end{bmatrix}, \begin{bmatrix} \frac{1}{2}(2v_{it}W^{-1}\xi_0 + v_{it} - x) \\ u_{it} + v_{it}\xi_0^\top W^{-1}\xi_0 - v_{it} \end{bmatrix} \geq 0,
\]
and thus, (A2) is equivalent to
\[
q_{it} \geq \min_{v_{it} \geq 0, u_{it} \in \mathbb{R}} u_{it} - \left[ (b')^\top x + h^i \right] + \alpha_t - v_{it}^\top \xi^i
\text{ s.t. } \begin{bmatrix} -\frac{1}{2}(2v_{it}W^{-1}\xi_0 + v_{it} - x)^\top \end{bmatrix}, \begin{bmatrix} -\frac{1}{2}(2v_{it}W^{-1}\xi_0 + v_{it} - x) \\ u_{it} + v_{it}\xi_0^\top W^{-1}\xi_0 - v_{it} \end{bmatrix} \geq 0.
\]
Then the claim follows.

**Proof of Proposition 3.4:** Note that (32b) can be rewritten as
\[
q_{it} \geq G_{fi} (v_{it}, 1, x) + \alpha_t - v_{it}^\top \xi^i
\]
\[
= \sup_{\xi \in \Xi} \left[ v_{it}^\top \xi - w_t(\xi)^\top x \right] + \alpha_t - v_{it}^\top \xi^i
\]
\[
= \sup_{\xi \in \Xi} \left[ v_{it}^\top \xi - \left( \xi^\top W_t(x)\xi + R_t(x)^\top \xi + H_t(x) \right) \right] + \alpha_t - v_{it}^\top \xi^i
\]
\[
= \sup_{\xi \in \Xi} \left[ (v_{it} - R_t(x))^\top \xi - \xi^\top W_t(x)\xi \right] - H_t(x) + \alpha_t - v_{it}^\top \xi^i.
\]
We now discuss the reformulation of the inner subproblem \( \sup_{\xi \in \Xi} \) in (A3). Similar to the proof of Proposition 3.3, for this problem, there exists \( \xi = \xi_0 \) such that \((\xi - \xi_0)^\top W^{-1}(\xi - \xi_0) < 1\), which implies Slater’s condition is satisfied, and hence the strong duality holds.

Thus, we have
\[
\sup_{\xi \in \Xi} \left( (v_{it} - R_t(x))^\top \xi - \xi^\top W_t(x)\xi \right) = \inf_{v_{it} \geq 0} \sup_{\xi \in \Xi} \left[ (v_{it} - R_t(x))^\top \xi - \xi^\top W_t(x)\xi - v_{it}((\xi - \xi_0)^\top W^{-1}(\xi - \xi_0) - 1) \right],
\]
which is equivalent to
\[
\min_{\nu_t \geq 0, u_{it} \in \mathbb{R}} u_{it}
\]
\[
s.t. \quad (v_{it} - R_t(x))^\top \xi - \xi^\top W_t(x) \xi - v_{it}((\xi - \xi_0)^\top W^{-1}(\xi - \xi_0) - 1) \leq u_{it}, \quad \forall \xi, \in \mathbb{R}^m,
\]
which can be further written as
\[
\min_{\nu_t \geq 0, u_{it} \in \mathbb{R}} u_{it}
\]
\[
s.t. \quad \left[ W_t(x) + v_{it} W^{-1} - \frac{1}{2} (2v_{it} W^{-1} \xi_0 + v_{it} - R_t(x))^\top u_{it} + v_{it} \xi_0^\top W^{-1} \xi_0 - v_{it} \right] \succeq 0,
\]
and thus, (A3) is equivalent to
\[
q_{it} \geq \min_{\nu_t \geq 0, u_{it} \in \mathbb{R}} u_{it} - H_t(x) + \alpha_t - v_{it}^\top \zeta^i
\]
\[
s.t. \quad \left[ -\frac{1}{2} (2v_{it} W^{-1} \xi_0 + v_{it} - R_t(x))^\top u_{it} + v_{it} \xi_0^\top W^{-1} \xi_0 - v_{it} \right] \succeq 0.
\]
Then the claim follows.