Data-Driven Distributionally Robust Chance-Constrained Optimization with Wasserstein Metric

Ran Ji
Department of Systems Engineering and Operations Research, George Mason University, rji2@gmu.edu;

Miguel A. Lejeune
Department of Decision Sciences, The George Washington University, mlejeune@gwu.edu

We study distributionally robust chance-constrained programming (**DRCCP**) optimization problems with data-driven Wasserstein ambiguity sets. The proposed algorithmic and reformulation framework applies to distributionally robust optimization problems subjected to individual as well as joint chance constraints, with random right-hand side and technology vector, and under two types of uncertainties, called uncertain probabilities and continuum of realizations. For the uncertain probabilities case, we provide new mixed-integer linear programming reformulations for **DRCCP** problems and derive a set of precedence optimality cuts to strengthen the formulations. For the continuum of realizations case with random right-hand side, we propose an exact mixed-integer second-order cone programming (MISOCP) reformulation and a linear programming (LP) outer approximation. For the continuum of realizations case with random technology vector, we propose two MISOCP and LP outer approximations. We show that all proposed relaxations become exact reformulations when the decision variables are binary or bounded general integers. We evaluate the scalability and tightness of the proposed MISOCP and (MI)LP formulations on a distributionally robust chance-constrained knapsack problem.

Key words: Distributionally Robust Optimization, Chance-Constrained Programming, Wasserstein Metric, Mixed-Integer Programming

History: This version: Feb 2020

---

1. Introduction

1.1. Problem Setting

Data-driven prescriptive analytics has been receiving growing attention along the emergence of data-centric environments. The continuously increasing availability of data provides great opportunities but also raises concerns revolving around data veracity IBM (2017), leading for example to question whether the available data can provide an exact characterization of the sources of uncertainty and their probability distribution. Distributionally robust optimization (see, e.g., Delage and Ye (2010) and Wiesemann et al. (2014)) has been shown to be a most suitable modeling paradigm to account for such uncertainty.

The objective of this paper is to investigate a class of data-driven distributionally robust chance-constrained programming models, in which the stochastic constraints are satisfied with a specified probability level across all possible probability distributions within an ambiguity set. The generic formulation of the distributionally robust chance-constrained problem **DRCCP** studied here is:

$$\textbf{DRCCP:} \quad \min_{x} \quad g(x) \tag{1a}$$

$$\text{s.t.} \quad \inf_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\left\{\xi : h(x,\xi) \geq 0\right\} \geq 1-\epsilon \tag{1b}$$

$$x \in \mathcal{X} \tag{1c}$$

where $x \in \mathbb{R}^M$ is a vector of decision variables, $\mathcal{X}$ is a feasible region defined by deterministic constraints, and $\epsilon \in (0,1)$ is a fixed probability level. The random variable ξ is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, in which Ω represents the sample space of the random variable ξ with a sigma algebra $\mathcal{F}$, and $\mathbb{P}$ denotes the associated probability measure supported on Ω. The **DRCCP** model seeks a solution $x \in \mathcal{X}$ that minimizes a convex cost function $g(x)$ while ensuring, via the ambiguous chance constraint (1b), that the worst-case probability of the event $\{h(x,\xi) \geq 0\}$ is at least equal to $1-\epsilon$ within the ambiguity set $\mathcal{D}$. Utilizing the complement rule, the ambiguous chance constraint (1b) is equivalent to:

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\left\{\xi : h(x,\xi) < 0\right\} \leq \epsilon \tag{2}$$

The equivalence between a chance constraint and an expectation one with indicator function allows the rewriting of (2) as:

$$\sup_{\mathbb{P}\in\mathcal{D}} \ \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{h(x,\xi)<0\}}\right] \leq \epsilon \ , \tag{3}$$

where $\mathbb{1}_{\{h(x,\xi)<0\}}$ denotes the indicator function taking value 1 if $h(x,\xi) < 0$ happens and 0 otherwise.

Throughout this paper, we focus on the **DRCCP** model with two *uncertainty types* considered (see Pflug and Pohl (2017)):

1. *Uncertain probabilities*: The underlying probability distribution has a finite support and its atoms are those of the reference distribution. The mass probability of each atom is allowed to vary and unknown.

2. *Continuum of realizations*: The random variable ξ can have a continuum (infinite number) of realizations.

The following is assumed regarding the structure of the uncertain vector ξ, the ambiguity set $\mathcal{D}$, and the set defined by the inequality $h(x,\xi) < 0$:

- The random vector ξ is supported on a bounded and convex set Ω.
- The ambiguity set is constructed via a data-driven approach based on the Wasserstein metric, i.e., the ambiguity set $\mathcal{D}$ is supposed to contain all the underlying probability distributions $\mathbb{P}$ belonging to a Wasserstein ball centered at the empirical distribution consisted of a finite number of data samples.
- The stochastic function $h(x,\xi)$ is affine in both x and ξ. Two *functional forms* for the stochastic inequalities defining $\{\xi : h(x,\xi) < 0\}$ are specified[1]:

[1] In an abuse of notation, we use the same symbol ξ in chance-constrained models involving either random right-hand side or random technology vector models, although ξ represents a scalar in the first case and a vector of dimension M in the second one.

1. *Linear random right-hand side*: $\{\xi : h(x,\xi) < 0\} = \{\xi : a^T x < \xi\}$, where $a \in \mathbb{R}^M$ is deterministic and $\xi \in \mathbb{R}$ is the random right-hand side.

2. *Linear random technology vector*: $\{\xi : h(x,\xi) < 0\} = \{\xi : \xi^T x < b\}$, where $\xi \in \mathbb{R}^M$ is the random technology vector.

1.2. Literature Review

Approaches based on moments and statistical distances are commonly used to build ambiguity sets in distributionally robust optimization (DRO). Moment-based ambiguity sets include the underlying distributions that have moment similarities (e.g., mean, variance) with the true probability distribution and often yield to computationally tractable conic formulations, see, e.g., Calafiore (2007), Delage and Ye (2010), Zymler et al. (2013), Cheng et al. (2014), Yang and Xu (2016), Hanasusanto et al. (2017) and the references therein.

We now proceed to review distributionally robust chance-constrained (DRCC) studies with statistical ambiguity sets considering the type of uncertainty (uncertain probability vs. continuum or realizations), the form of the random inequalities (linear vs. nonlinear), and the type of the obtained formulation (reformulation vs. approximation). We give particular attention to studies based on the Wasserstein metric.

Erdoğan and Iyengar (2006) consider a linear random technology vector and construct an ambiguity set based on the Prohorov-metric. They obtain an approximated robust solution from a sampling problem, where each constraint in the ambiguity set corresponds to one sample data drawn from the reference distribution. Jiang et al. (2016) investigate the distributionally robust chance-constrained unit commitment problem with renewable energy with an ambiguity set in which the L_2 norm distance between the mass probabilities of the underlying and reference distributions is bounded from above. They study the uncertain probability case in an individual linear chance constraint with random right-hand side, reformulate the problem as a mixed-integer second order cone problem, and derive valid inequalities. Considering joint distributionally robust chance constraints, Jiang and Guan (2015) and Hu and Hong (2013) derive tractable reformulations for the family of ϕ-diverge probability metric (e.g., Kullback-Leibler (KL) divergence, Hellinger distance, etc.). These studies show that a distributionally robust chance constraint is equivalent to a classical chance constraint under the reference distribution with a re-scaled reliability level, and cover the cases of linear uncertainty in the right-hand side and technology matrix.

A reason for the widespread use of Wasserstein ambiguity sets in DRO is that they enjoy the finite sample guarantee property Esfahani and Kuhn (2018), which signifies that the Wasserstein ambiguity set can be constructed to contain the true probability distribution with a high probability level and can provide an upper bound on the out-of-sample performance of the distributionally robust solution. DRO studies on Wasserstein ambiguity sets can be categorized, as noted by Pflug and Pohl (2017), with respect to the types of uncertainties presented in Section 1.1. See Pflug and Wozabal (2007), Postek et al. (2016), Ji and Lejeune (2017) for the uncertain probability case and Pflug et al. (2012), Zhao and Guan (2018), Gao and

Kleywegt (2016), Esfahani and Kuhn (2018) for the continuum of realizations uncertainty type, in which the underlying probability distribution can take infinitely many possible values.

Wasserstein ambiguity sets have been primarily used to study DRO expectation constraints. For example, in the uncertain probability case, Ji and Lejeune (2017) seek to maximize the worst-case expected fractional functions representing reward-risk ratios. The proposed reformulations derive the support function of Wasserstein ambiguity set and the concave conjugate of the ratio functions. As for the continuum of realization case, Esfahani and Kuhn (2018) utilize convex duality to derive reformulations of the worst-case expectation of a loss function representable as pointwise maxima. Blanchet et al. (2018) study the distributionally robust mean-variance portfolio optimization problem with Wasserstein ambiguity by reformulating the problem into an empirical variance minimization problem with a regularization term.

A few recent studies on distributionally robust chance constraints with Wasserstein ambiguity sets tackle problems with linear chance constraints in the continuum of realizations case. Duan et al. (2017) study a distributionally robust chance-constrained optimal power flow problem with Wasserstein ambiguity set. They construct a deterministic robust constraint to approximate the distributionally robust individual chance constraint and utilize a bisection search method to find the worst-case expectation of a quadratic cost function. Xie and Ahmed (2018) propose a bicriteria approximation scheme to solve a specific type of distributionally robust chance-constrained covering problems with Wasserstein ambiguity sets. This approach guarantees to find a solution that violates the stochastic inequalities with a probability that is upper-bounded by the original probability level multiplied by a constant called violation ratio. While the paper was under review, we became aware of two independent studies by Xie (2018) and Chen et al. (2018) closely related to ours. Both studies propose reformulation and approximations for **DRCCP** with Wasserstein ambiguity set under the continuum of realizations case which implicitly assumes that all atoms in the constructed worst-case distribution are equally likely, but they do not consider the case of uncertain probabilities. Our study is in that respect the most encompassing in the sense that it covers distributionally robust joint and individual chance constraints, with random right-hand side and random technology matrix, and in the uncertainty cases characterized by uncertain probabilities and continuum of realizations.

Xie (2018) uses the dual form of the worst-case expectation constraint with indicator function to reformulate the distributionally robust chance constraint with Wasserstein metric. The resulting reformulation admits a Conditional Value-at-Risk (CVaR) representation. Xie applies his approach to single and joint chance constraints with linear uncertainties in the right-hand side or in the technology matrix. He shows that **DRCCP** is mixed-integer representable by introducing big-M parameters and additional binary variables. He also proposes an outer approximation with a Value-at-Risk constraint and three inner approximation methods. For **DRCCP**s with binary decision variables, he develops a set of reformulations derived by exploiting the submodular structure of the problem. Chen et al. (2018) reach the same CVaR representation as Xie (2018) using a different proof scheme that utilizes the structural information of the worst-case probability distribution. They provide exact mixed-integer conic reformulations for problems with individual and joint chance constraints with random right-hand side and for individual chance constraints with random technology matrix. When the Wasserstein ambiguity set is constructed via l_1-norm or 1_∞ norm, their proposed mixed-integer conic program reduces to a mixed-integer linear program (MILP). They also show that approximation methods based on CVaR and Bonferroni inequalities do not perform well. In the continuum of realizations case studied by Xie (2018) and Chen et al. (2018), the early stage of our reformulation framework is similar to Xie (2018) and Chen et al. (2018) by using the worst-case expectation of the indicator function to represent the chance constraints. Our approaches differ from theirs in the later stages, when we rely on a second-order cone reformulation approach and on the McCormick envelop to tackle the bilinear terms involved in the reformulation. This allows us to provide exact reformulations and approximations taking the form of mixed-integer second-order cone programming (MISOCP) and (mixed-integer) linear programming problems.

1.3. Contributions, Structure, and Notations

We derive new tractable reformulations for several variants of the **DRCCP** problem with Wasserstein ambiguity under different (i) types of uncertainties (i.e., uncertain probabilities vs. continuum of realizations), and (ii) forms of stochastic inequalities (i.e., right-hand side vs. technology vector). We rely on conic duality theory and the dual form of the worst-case expectation of an indicator function to subsequently derive mixed-integer linear or conic reformulations/approximations for the distributionally robust chance constraint. In the uncertain probability case, we provide mixed-integer linear programming reformulations for **DRCCP** problems with random right-hand side and technology vector. We also derive two sets of precedence valid inequalities that strengthen the formulations. In the continuum of realizations case, we propose an exact mixed-integer second-order cone programming (MISOCP) reformulation and a linear programming (LP) relaxation for chance constraints with random right-hand side, while, for chance constraints with random technology vector, we propose two MISOCP and LP relaxations. All of our proposed MISOCP or LP relaxations are shown to be equivalent to the true formulation when the decision variables are bounded integers. We carry out numerical experiments on instances of the distributionally robust chance-constrained knapsack problems to show the effectiveness of the proposed MISOCP and (MI)LP reformulation/approximation framework.

The rest of the paper proceeds as follows. Section 2 sketches the construction of Wasserstein ambiguity sets. Section 3 and 4 derive the reformulation and approximation methods for the **DRCCP** problems with individual chance constraints in the uncertain probabilities and continuum of realizations cases. Section 5 extends the solution framework for individual chance constraints to joint chance constraints. The results of the computational experiments are reported in Section 6. Section 7 concludes.

Notation. We denote by $\mathbb{R}_+$ and $\mathbb{R}_{++}$ the set of non-negative and positive real numbers, respectively. The round-up and round-down of a real number a are respectively $\lceil a \rceil$ and $\lfloor a \rfloor$ and e denotes the unit vector. Let $\mathcal{D}$ be the ambiguity set, $\mathbb{P} \in \mathcal{D}$ be an underlying probability distribution, $\mathbb{P}_0$ be the reference (empirical) distribution, and $\mathbb{Q}$ represent the true and unknown probability distribution. Let $p_j = P(\xi = \xi_j), \forall j \in \mathcal{N}$ denote the probability associated with realization j and $\mathcal{P}(\Omega)$ represent the space of all probability distributions $\mathbb{P}$ supported on Ω. We denote by δ_ξ the Dirac distribution concentrating unit mass at ξ. We define the characteristic function as $\chi_\Omega(\xi) = 0$, if $\xi \in \Omega$; $= \infty$, otherwise. We use the index sets $\mathcal{N} = \{1, \ldots, N\}$, $\mathcal{M} = \{1, \ldots, M\}$. For a given norm $||\cdot||$ on $\mathbb{R}^M$, the dual norm $||\cdot||_*$ is defined by $||v||_* = \sup\{v^T\zeta : ||\zeta|| \leq 1\}$. The conjugate of a real-valued function $f(v)$ is defined as $f^*(v) = \sup_\zeta v^T\zeta - f(\zeta)$. The Wasserstein ambiguity set is denoted by $\mathcal{D}_W^{UP}$ for the uncertain probability case and $\mathcal{D}_W^{CR}$ for the continuum of realizations one.

2. Preliminaries on Wasserstein Ambiguity Set

To be self-contained, we briefly present some key concepts involved in the construction of Wasserstein ambiguity sets. The Wasserstein metric between two probability distributions is defined as the minimum mass transportation cost incurred by moving one distribution to another, where the unit cost is defined as the norm distance between two atoms[2]. Throughout this paper unless stated otherwise, we restrict our attention to the 1-Wasserstein distance based on the l_1 norm.

DEFINITION 1. (Wasserstein Metric) The Wasserstein metric $d_W(\mathbb{P}, \mathbb{P}_0) : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \to \mathbb{R}$ between two probability distributions $\mathbb{P}, \mathbb{P}_0 \in \mathcal{P}(\Omega)$ is defined by:

$$d_W(\mathbb{P}, \mathbb{P}_0) = \inf \left(\int_{\Omega^2} ||\xi - \xi^0|| \Pi(d\xi, d\xi^0) : \begin{array}{c} \Pi \text{ is a joint distribution of } \xi, \\ \xi^0 \text{ with marginals } \mathbb{P} \text{ and } \mathbb{P}_0 \end{array} \right) . \quad (4)$$

For a discrete distribution problem, the Wasserstein metric is given by:

$$d_W(\mathbb{P}, \mathbb{P}_0) = \inf_{\pi \geq 0} \left(\sum_{i,j \in \mathcal{N}} \pi_{ij} ||\xi_j - \xi_i^0|| : \begin{array}{l} \sum_{j \in \mathcal{N}} \pi_{ij} = p_i^0, \ \forall i \in \mathcal{N} \\ \sum_{i \in \mathcal{N}} \pi_{ij} = p_j, \ \forall j \in \mathcal{N} \end{array} \right) , \quad (5)$$

where π_{ij} represents the probability mass transported from ξ_i^0 to ξ_j, p_i^0 and p_j respectively denote the probability of atom ξ_i^0 and of ξ_j.

The Wasserstein ambiguity set $\mathcal{D}_W$ is a ball of radius θ centered around the empirical distribution $\mathbb{P}_0$

$$\mathcal{D}_W := \{\mathbb{P} \in \mathcal{P}(\Omega) : d_W(\mathbb{P}, \mathbb{P}_0) \leq \theta\} , \quad (6)$$

where $\mathbb{P}_0$ is constructed with N i.i.d samples: $\mathbb{P}_0 := \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_i^0}$ and $\delta_{\xi_i^0}$ refers to the Dirac distribution concentrating unit mass at realization ξ_i^0. For the selection of the radius θ of Wasserstein ambiguity set, we refer to Ji and Lejeune (2017) for more detailed discussions.

[2] We shall interchangeably use the terms atoms and realizations of uncertain variables ξ.

As previously demonstrated (Gao and Kleywegt 2016, Theorem 1), strong duality holds true when the Wasserstein ambiguity set is constructed via a data-driven approach as in (6), which allows the rewriting of the worst case expectation $\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{h(x,\xi)<0\}}\right]$ as:

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{h(x,\xi)<0\}}\right] = \inf_{\lambda\geq 0}\left[\lambda\theta + \sup_{\xi\in\Omega}\mathbb{E}\left[\mathbb{1}_{\{h(x,\xi)<0\}} - \lambda d_W(\xi,\xi^0)\right]\right]. \tag{7}$$

We utilize this duality result to derive exact reformulations or approximate formulations for the two uncertainty types: uncertain probabilities in Section 3 and continuum of realizations in Section 4.

3. Wasserstein DRCCP – Case with Uncertain Probabilities

We first focus on the case with *uncertain probabilities* in which the distance between atoms is a constant. To ease the notations, we set $c_{ij} = ||\xi_j - \xi_i^0||, \forall i,j\in\mathcal{N}: i\neq j$ to denote the unit cost of transporting the probability mass from atom ξ_i^0 to another one ξ_j. We recall that: 1) the probabilities of the atoms of the underlying probability distribution are allowed to vary and are defined as decision variables and 2) the atoms of the underlying distribution are known and identical to those of the empirical distribution.

We denote by $\mathbb{1}_{\{h(x,\xi_j)<0\}}$ the indicator function when ξ is realized as ξ_j which occurs with probability p_j. Let p_i^0 be the probability of the i^{th} atom in the empirical distribution and $\mathcal{D}_W^{UP}$ be the Wasserstein ambiguity set in the case with uncertain probabilities. Theorem 1 reformulates the worst-case expectation semi-infinite programming problem $\sup_{\mathbb{P}\in\mathcal{D}_W^{UP}} \mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{h(x,\xi)<0\}}]$ in a finite dimensional space.

THEOREM 1. *(DRCCP Reformulation for Uncertain Probabilities) Let $w\in\{0,1\}^N$. For any $\theta\geq 0$, the distributionally robust chance constraint* (1b) *with uncertain probabilities can be reformulated with the set of mixed-integer linear inequalities $\mathcal{Z}^{UP}$*

$$(x,\lambda,v,w)\in\mathcal{Z}^{UP} = \begin{cases} \lambda\theta + \sum_{i\in\mathcal{N}} p_i^0 v_i \leq \epsilon & \text{(8a)}\\ \lambda\geq 0, & \text{(8b)}\\ \lambda c_{ij} + v_i - w_j \geq 0, \quad \forall i,j\in\mathcal{N}: i\neq j & \text{(8c)}\\ w_j = \mathbb{1}_{\{h(x,\xi_j)<0\}}, \quad \forall j\in\mathcal{N} & \text{(8d)} \end{cases}$$

Proof. The value of the worst-case expectation can be obtained from:

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{D}_W^{UP}} \mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{h(x,\xi)<0\}}] = \sup_{\pi,p\geq 0} \quad & \sum_{j\in\mathcal{N}} p_j \mathbb{1}_{\{h(x,\xi_j)<0\}} && \text{(9a)}\\ s.t. \quad & \sum_{i,j\in\mathcal{N}} \pi_{ij}c_{ij} \leq \theta && \text{(9b)}\\ & \sum_{j\in\mathcal{N}} \pi_{ij} = p_i^0, \quad \forall i\in\mathcal{N} && \text{(9c)}\\ & \sum_{i\in\mathcal{N}} \pi_{ij} = p_j, \quad \forall j\in\mathcal{N} && \text{(9d)} \end{aligned}$$

Using the equality constraints (9d), we first remove p_j from (9a) and from problem (9) altogether. Denoting by $\lambda \geq 0$ and $v_i, i \in \mathcal{N}$ the dual variables corresponding to (9b) and (9c), respectively, we use Lagrangian duality to reformulate problem (9) as:

$$\sup_{\mathbb{P}\in\mathcal{D}_W^{UP}} \mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{h(x,\xi)<0\}}] = \sup_{\pi\geq 0} \inf_{\lambda\geq 0, v} \sum_{i,j\in\mathcal{N}} \pi_{ij} \mathbb{1}_{\left\{h(x,\xi_j)<0\right\}} + \lambda\left(\theta - \sum_{i,j\in\mathcal{N}} \pi_{ij} c_{ij}\right) + \sum_{i\in\mathcal{N}} v_i \left(p_i^0 - \sum_{j\in\mathcal{N}} \pi_{ij}\right) \tag{10a}$$

$$= \inf_{\lambda\geq 0, v} \sup_{\pi\geq 0} \lambda\theta + \sum_{i\in\mathcal{N}} p_i^0 v_i - \sum_{i,j\in\mathcal{N}} \pi_{ij}\left(\lambda c_{ij} + v_i - \mathbb{1}_{\left\{h(x,\xi_j)<0\right\}}\right) \tag{10b}$$

$$= \inf_{\lambda\geq 0, v} \lambda\theta + \sum_{i\in\mathcal{N}} p_i^0 v_i + \sup_{\pi\geq 0}\left(-\sum_{i,j\in\mathcal{N}} \pi_{ij}\left(\lambda c_{ij} + v_i - \mathbb{1}_{\left\{h(x,\xi_j)<0\right\}}\right)\right) \tag{10c}$$

$$= \inf_{\lambda\geq 0, v} \lambda\theta + \sum_{i\in\mathcal{N}} p_i^0 v_i + \begin{cases} 0, & \text{if } \lambda c_{ij} + v_i - \mathbb{1}_{\left\{h(x,\xi_j)<0\right\}} \geq 0, \forall i,j\in\mathcal{N} \\ +\infty, & \text{otherwise} \end{cases} . \tag{10d}$$

Equations (10b) and (10c) are obtained by rearranging the terms in (10a) while (10d) gives an the upper bound for $(-\sum_{i,j\in\mathcal{N}} \pi_{ij}\left(\lambda c_{ij} + v_i - \mathbb{1}_{\left\{h(x,\xi_j)<0\right\}}\right))$ by maximizing over π. Using (10d) to replace the worst-case expectation $\sup_{\mathbb{P}\in\mathcal{D}_W^{UP}} \mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{h(x,\xi)<0\}}]$ into (3) gives (8a) and (8c). Introducing the linkage between w_j and $\mathbb{1}_{\left\{h(x,\xi_j)<0\right\}}$ leads to (8d). □

We next propose precedence optimality cuts on the binary variables w_j. The optimality cuts are derived from the set of constraints (8c) and the ordering of the cost coefficients $c_{i,j}$ defining the distance between atoms ξ_i^0 and ξ_j.

PROPOSITION 1. *(Precedence Optimality Cuts) Assume without loss of generality that $c_{i,j(i_k)}$ is the ordered vector of distances c_{ij} for atom $i\in\mathcal{N}$:*

$$c_{i,j(i_1)} \leq c_{i,j(i_2)} \leq \ldots \leq c_{i,j(i_k)} \leq \ldots \leq c_{i,j(i_N)}, \ \forall i\in\mathcal{N}, \tag{11}$$

where the index $j(i_k)$ identifies the atom ξ_j that has k-th smallest distance to ξ_i^0. The inequalities

$$w_{j(i_{k-1})} \leq w_{j(i_k)}, \quad \forall i, j(i_k)\in\mathcal{N} : k = 2,\ldots,N, i\neq j(i_k) \tag{12}$$

are optimality cuts for $\mathcal{Z}^{UP}$.

Proof. Using the above notation for the ordered vector of distances, we rewrite (8c) as

$$\lambda c_{i,j(i_k)} + v_i - w_{j(i_k)} \geq 0, \quad \forall i, j(i_k)\in\mathcal{N} : i\neq j(i_k). \tag{13}$$

We distinguish two cases depending on the value of λ:

(i) $\lambda > 0$: For any given $i\in\mathcal{N}$ with fixed values of λ and v_i, due to the ascending ordering of distances defined by (11) and constraints (13), it follows that, if $w_{j(i_k)} = 0$, $w_{j(i_{k-1})}$ must take value of 0 for $k\geq 2$.

This validates the precedence constraints (12) when $\lambda > 0$.

(ii) $\lambda = 0$: The constraints (8c) reduce to

$$v_i - w_{j(i_k)} \geq 0, \quad \forall i, j(i_k) \in \mathcal{N} : i \neq j(i_k) \ . \tag{14}$$

(ii.1) If $v_i < 0$, the feasible set is empty.

(ii.2) If $0 \leq v_i < 1$, the proof is the same as in (i).

(ii.3) Consider now the case $v_i \geq 1$ in which the constraints (14) do not impose any restriction on the variables $w_{j(i_k)}$: all the 0-1 components of the N-dimensional cube $\{0,1\}^N$ are feasible. It can easily be seen that it is sufficient to only consider the solution in which all components of the vector w are equal to 0 (thereafter denoted w^0), since the feasible set associated with the solution vector w^0 strictly includes the feasible set of all the other solutions in which one or more components of w has a non-zero (i.e., 1) value. Indeed, the feasible set $\mathcal{Z}^{UP}$ associated with w^0 is the set $\mathcal{Z}^{UP^0}$ defined below:

$$(x,\lambda,v,w) \in \mathcal{Z}^{UP^0} \begin{cases} \sum_{i \in \mathcal{N}} p_i^0 v_i \leq \epsilon & \text{(15a)} \\ \lambda = 0 \geq 0, & \text{(15b)} \\ v_i - \underbrace{0}_{w_j^0} \geq 0, \quad \forall i,j \in \mathcal{N} : i \neq j & \text{(15c)} \end{cases}$$

On the other hand, the feasible set $\mathcal{Z}^{UP}$ of any other feasible integer solution $w' \in \{0,1\}^N, w' \neq w^0$ (in which at least one component of w' takes value 1) is the set $\mathcal{Z}^{UP'}$:

$$(x,\lambda,v,w) \in \mathcal{Z}^{UP'} \begin{cases} \sum_{i \in \mathcal{N}} p_i^0 v_i \leq \epsilon & \text{(16a)} \\ \lambda = 0 \geq 0, & \text{(16b)} \\ v_i - \underbrace{1}_{w'_j} \geq 0, \quad \forall i \in \mathcal{N}, j \in \mathcal{N}^1 : i \neq j & \text{(16c)} \\ v_i - \underbrace{0}_{w'_j} \geq 0, \quad \forall i \in \mathcal{N}, j \setminus \mathcal{N}^1 : i \neq j & \text{(16d)} \end{cases}$$

with $\mathcal{N}^1$ denoting the subset of components w'_j of w' taking value 1.

Comparing (15c) on one hand with (16c) and (16d) on the other hand, it is straightforward to see that $\mathcal{Z}^{UP^0} \supset \mathcal{Z}^{UP'}$. Therefore, considering w^0 only and optimizing accordingly over $\mathcal{Z}^{UP^0}$ is sufficient to reach the optimal solution, which in turn implies that cutting off any of the feasible solutions w' will not preclude from reaching the optimal value of the problem.

The precedence inequalities (12) just do that and cut off all vectors $w' \in \{0,1\}^N$ in which there is at least one pair $(w'_j, w'_{j+r}), r \geq 1, (j+r) \in \mathcal{N}$ of components such that: $(w'_j, w'_{j+r}) = (1,0)$. Therefore, the precedence inequalities (12) cut off integer feasible solutions, but do not cut any feasible integer solution that will prevent from finding the optimal value of the problem. Hence, they qualify as optimality cuts, which is the result we set out to prove. □

When $\lambda > 0$, the precedence constraints (12) do not cut any integer solution feasible for the system (8a)-(8d), but cut fractional solutions that would be feasible for the continuous relaxation of the set of mixed-integer linear inequalities (8a)-(8d). The precedence constraints (12) can be added to the formulation to tighten the continuous relaxation as illustrated by the following example. Consider a case with three data points and distances $c_{12} = 1, c_{13} = 2$. Let $i = 1$, $\lambda = 0.5$, and $v_1 = -0.2$. The constraints (8c) read $w_2 \leq \lambda \cdot 1 + v_1 = 0.3$ and $w_3 \leq \lambda \cdot 2 + v_1 = 0.8$, which means that $(w_2, w_3) = (0.25, 0.15)$ is a feasible solution for the continuous relaxation of the system (8a)-(8d). However, adding (12) imposes $w_2 \leq w_3$ and cuts this fractional solution, thereby tightening the feasible area of the continuous relaxation of (8a)-(8d). When $\lambda = 0$, the precedence constraints (12) cut some integer feasible solutions without preventing the finding of the optimal value.

In the next subsections, we consider two functional forms for the stochastic inequalities defining the set $\{\xi : h(x, \xi) < 0\}$ and define for each the specific formulation of the set $\mathcal{Z}^{UP}$ in Theorem 1. Note that the precedence cuts (11)-(12) derived above are valid for the two considered functional forms (i.e., random right-hand side in Section 3.1 and random technology vector in Section 3.2) and can be added to strengthen each of the formulations presented in the next two subsections.

3.1. Random Right-Hand Side

We now consider stochastic inequalities with linear random right-hand side. The set $\{\xi : h(x, \xi) < 0\}$ is defined by $\{\xi : a^T x < \xi\}$ with $\xi \in \mathbb{R}$. Corollary 1 follows from Theorem 1.

COROLLARY 1. *(Linear Uncertainty with Random Right-Hand Side) Let L_x, $U_x \in \mathbb{R}^M$ be the vectors of lower and upper bounds for x, respectively. Let $\varrho > 0$ be an infinitesimal positive number and S_j be a positive constant defined as:*

$$S_j = \sum_{m \in \mathcal{M}: a_m > 0} a_m U_{x_m} + \sum_{m \in \mathcal{M}: a_m \leq 0} a_m L_{x_m} - \xi_j, \quad j \in \mathcal{N}.$$

The distributionally robust chance constraint

$$\sup_{\mathbb{P} \in \mathcal{D}_W^{UP}} \mathbb{P}\left\{\xi : a^T x < \xi\right\} \leq \epsilon \tag{17}$$

can be reformulated with the set of mixed-integer linear inequalities $\mathcal{Z}_R^{UP}$:

$$(x, \lambda, v, w) \in \mathcal{Z}_R^{UP} = \begin{cases} a^T x \leq \xi_j + S_j(1 - w_j) - \varrho w_j, \quad \forall j \in \mathcal{N} & \text{(18a)} \\ w_j \in \{0, 1\}, \quad \forall j \in \mathcal{N} & \text{(18b)} \\ \text{(8a)} - \text{(8c)} \ . \end{cases}$$

In the uncertain probability case, the mixed-integer linear programming problem

$$\min_{x, \lambda, v, w} \left\{ g(x) : (x, \lambda, v, w) \in \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^N \times \{0, 1\}^N : (x, \lambda, v, w) \in \mathcal{X} \cap \mathcal{Z}_R^{UP} \right\} \tag{19}$$

is equivalent to **DRCCP** *for linear inequalities with random right-hand side.*

Proof. Theorem 1 allows the reformulation of (17) with the set of constraints (8a)-(8d), in which (8d) can then be replaced by (18a) and (18b). Each inequality (18a) ensures that the corresponding binary variable w_j takes value 0 if $a^T x \geq \xi_j$. □

We derive in Theorem 2 a tightened formulation for the set $\mathcal{Z}_R^{UP}$ in Corollary 1. Let ϱ be an infinitesimally small number.

THEOREM 2. *(Strengthened Formulation) Assume without loss of generality that the $\xi_j, j = 1, \ldots, N$ are indexed in such a way that*

$$\xi_1 \leq \xi_2 \leq \ldots \leq \xi_N \; . \tag{20}$$

The set $\mathcal{Z}_R^{UP}$ can be reformulated and tightened as follows:

$$(x, \lambda, v, w, \pi) \in \mathcal{Z}_{RT}^{UP} \begin{cases} w_{j-1} \leq w_j \;,\; j = 2, \ldots, N & \text{(21a)} \\ w_j - \displaystyle\sum_{k=1}^{j} \sum_{i \in \mathcal{N}} \pi_{ik} + \varrho \leq \epsilon \;,\; j \in \mathcal{N} & \text{(21b)} \\ \text{(8a)} - \text{(8c)}; \text{(18a)} - \text{(18b)} \end{cases}$$

Proof. Due to the ascending ordering (20) of the atoms and constraints (18a), it follows that 1) for $j \geq 2$, if w_j must take value 0, then $w_{j-\ell}, \ell = 1, \ldots, j$ cannot take value 1, and 2) if $w_j = 1$, then w_{j+1} can also take value 1 for $j \geq 1$. Both conditions are enforced by the precedence constraints (21a), which validates them.

We have now to demonstrate that (21b) are valid for $\mathcal{Z}_R^{UP}$. The probability of atom ξ_j is $p_j = \sum_{i \in \mathcal{N}} \pi_{ij}$ (see (9d)), and it ensures that the cumulative probability of atoms ξ_j is $\sum_{k=1}^{j} \sum_{i \in \mathcal{N}} \pi_{ik}$ due to (20). For the chance constraint (1b) to hold with probability at most ϵ, the binary variable w_j associated to each atom whose cumulative probability is equal or below ϵ must take value 0, which is enforced by (21b). The inequalities (21b) do not impose any restriction on the variables w_j associated with atoms with cumulative probability $\sum_{k=1}^{j} \sum_{i \in \mathcal{N}} \pi_{ik} > 1 - \epsilon$ and are thus valid. □

We emphasize here the need to introduce binary variables in the formulations presented in Corollary 1 and Theorem 2. Using the atoms' ordering defined by (20) can at first glance give the impression that the individual chance constraint (17) can be reformulated without binary variables and with one linear inequality $a^T x < \zeta^* = \xi_{j^*} - \varrho$, where ζ^* is the smallest ζ that satisfies the distributionally robust chance constraint $\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\xi : \zeta < \xi\} \leq \epsilon$, and ξ_{j^*} is the atom such that j^* is the smallest index value for which $\sum_{j=1}^{j^*} p_j \geq 1 - \epsilon$ and ϱ is a very small positive number. One may argue that the optimal value of ζ^* can be found independently from the decision variable x and without solving the reformulated distributionally robust problem. This would be true with traditional individual chance constraints having random right-hand side, in which setting the probabilities p_j are known and the value of j^* and ξ_{j^*} can be determined prior optimizing. In this case, ξ_{j^*} can be defined as a fixed parameter and $a^T x \leq \xi_{j^*} - \varrho$ is a linear (continuous) inequality. However, in the DRO setting, the value of ζ can only be found from the worst-case probability distribution $\mathbb{P}$ of ξ contained in the ambiguity set $\mathcal{D}$. To be more specific, in the **DRCCP** problem with uncertain probabilities case, the probabilities of the atoms defining the worst-case probability distribution are unknown a priori, and they are regarded as decision variables whose values are uncovered by solving the optimization problem. The optimal value of ζ cannot be found prior to solving the DRO problem and without knowing the worst-case probability distribution. The value of ζ^* is indeed not dependent on x, but depends on other decision variables, i.e., the uncertain probabilities p_j. The optimal value of ζ^* thus cannot be fixed a priori, but has to be uncovered by solving the reformulated DRO problem. In sum, the uncertain probabilities p_j is a decision variable and the finding of its value requires the introduction of the binary variables w_j in Corollary 1 and Theorem 2's formulations. The binary variables are also needed to ensure that the distance between the derived worst-case and the given nominal distributions does not exceed the radius of the Wassertsein ambiguity ball.

3.2. Random Technology Vector

We now consider a linear stochastic inequality with uncertainty in the technology vector. The set $\{\xi : h(x,\xi) < 0\}$ takes the form $\{\xi : \xi^T x < b\}$ with $\xi \in \mathbb{R}^M$.

COROLLARY 2. *(Linear Uncertainty in Random Technology Vector) Let $\varrho' > 0$ be an infinitesimal positive number and S'_j be a positive number:*

$$S'_j = \sum_{m \in \mathcal{M}:\xi_{jm} > 0} \xi_{jm} U_{x_m} + \sum_{m \in \mathcal{M}:\xi_{jm} \leq 0} \xi_{jm} L_{x_m} - b, \quad j \in \mathcal{N}.$$

The distributionally robust chance constraint

$$\sup_{\mathbb{P} \in \mathcal{D}_W^{UP}} \mathbb{P}\left\{\xi : \xi^T x < b\right\} \leq \epsilon \tag{22}$$

can be reformulated with the following set of mixed-integer linear inequalities $\mathcal{Z}_L^{UP}$:

$$(x,\lambda,v,w) \in \mathcal{Z}_L^{UP} = \begin{cases} \xi_j^T x \leq b + S'_j(1-w_j) - \varrho' w_j, \quad \forall j \in \mathcal{N} \\ \text{(8a)} - \text{(8c); (18b)} \end{cases} \tag{23a}$$

In the uncertain probability case, the mixed-integer linear programming problem

$$\min_{x,\lambda,v,w} \left\{ g(x) : (x,\lambda,v,w) \in \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^N \times \{0,1\}^N : (x,\lambda,v,w) \in \mathcal{X} \cap \mathcal{Z}_L^{UP} \right\} \tag{24}$$

is equivalent to **DRCCP** *for linear uncertainty in the technology vector.*

Proof. Theorem 1 allows the reformulation of (22) with the inequalities (8a)-(8d). The constraints (8d) can then be replaced by (18b) and (23a). Each inequality (23a) ensures that the corresponding binary variable w_j takes value 0 if $\xi_j^T x \geq b$. □

4. Wasserstein DRCCP – Case with Continuum of Realizations

We now consider the case when the random variable ξ can have a *continuum of realizations* (i.e, ξ can be varied to uncountably many possible values). The atoms ξ of the underlying distribution are unknown, instead of the probabilities p which were the unknowns in the uncertain probability case (Section 3). We propose exact reformulations for chance-constrained models with random right-hand side and outer approximations for those with random technology vector. We also define under which conditions the true and the approximation problems are equivalent.

We first recall the approach proposed by Esfahani and Kuhn (2018) to reformulate the worst-case expectation problem introduced earlier.

THEOREM 3. *(Dual Reformulation for Continuum of Realizations Esfahani and Kuhn (2018)) Let* $l(x,\xi) := \max_{k\in\mathcal{K}} l^{(k)}(x,\xi)$ *represent a pointwise maximum function, and the negative constituent functions* $-l^{(k)}(x,\xi), \forall k \in \mathcal{K}$ *are proper, convex and lower semi-continuous. Let* $s_i, \forall i \in \mathcal{N}$ *be a set of auxiliary variables. For any given* $\theta \geq 0$*, the worst-case expectation* $\sup_{\mathbb{P}\in\mathcal{D}_W^{CR}} \mathbb{E}_{\mathbb{P}}[l(x,\xi)]$ *with continuum of realizations is equal to the optimal value of the following problem:*

$$\inf_{\lambda,s,z,v} \quad \lambda\theta + \frac{1}{N}\sum_{i=1}^{N} s_i \tag{25a}$$

$$\text{s.t.} \quad [-l^{(k)}]^*(z_{ik} - v_{ik}) + \sigma_\Omega(v_{ik}) - z_{ik}^T \xi_i^0 \leq s_i \qquad \forall i \in \mathcal{N}, k \in \mathcal{K} \tag{25b}$$

$$||z_{ik}||_* \leq \lambda \qquad \forall i \in \mathcal{N}, k \in \mathcal{K} \tag{25c}$$

$$\lambda \geq 0 \tag{25d}$$

where x*,* λ*,* z*,* v *and* s *are decision variables,* $\lambda \geq 0$ *is the dual variable for the Wasserstein distance constraint* $\int_{\Omega^2} ||\xi - \xi^0|| \Pi(d\xi, d\xi^0) \leq \theta$*,* $[-l^{(k)}]^*(z_{ik} - v_{ik})$ *is the conjugate of* $-l^{(k)}(x,\xi)$ *at* $z_{ik} - v_{ik}$*,* $||z_{ik}||_*$ *is the dual norm, and* σ_Ω *is the support function of the uncertainty set* Ω*.*

We refer the reader to Esfahani and Kuhn (2018) for the proof of Theorem 3.

Lemma 1 represents the indicator function as a pointwise maxima. This will allow us to employ Theorem 3 to subsequently reformulate the distributionally robust chance constraint.

LEMMA 1. *The indicator function* $\mathbb{1}_{\{h(x,\xi)<0\}}$ *can be rewritten as the pointwise maximum of a finite number of concave functions, which is defined as*

$$l(x,\xi_i) = \max\{\underbrace{1 - \chi_{\{h(x,\xi_i)<0\}}}_{l_i^{(1)}}, \underbrace{0}_{l_i^{(2)}}\} \tag{26}$$

in which

$$\chi_{\{h(x,\xi_i)<0\}}(\xi) = \begin{cases} 0, \text{ if } h(x,\xi_i) < 0 \\ \infty, \text{ otherwise} \end{cases} \tag{27}$$

is the characteristic function of the convex set defined by $h(x,\xi_i) < 0$*.*

Proof. We first utilize the complement rule to represent the indicator function $\mathbb{1}_{\{h(x,\xi)<0\}}$: From (27), the equivalence between $\mathbb{1}_{\{h(x,\xi_i)<0\}}$ and (26) is immediate:

$$\max\{1-\chi_{\{h(x,\xi_i)<0\}}(\xi),0\} = \begin{cases} 1, \text{ if } \ h(x,\xi)<0 \\ 0, \text{ otherwise} \end{cases} \tag{28}$$

Since the characteristic function $\chi_{\{h(x,\xi_i)<0\}}$ of a convex set is convex (see Boyd and Vandenberghe (2004)), $l_i^{(1)}$ and $l_i^{(2)}$ are concave, and $l(x,\xi_i) = \max\{l_i^{(1)}, l_i^{(2)}\}$ represents a finite maxima of concave functions. □

The above representation of the indicator function $\mathbb{1}_{\{h(x,\xi)<0\}}$ is now used to reformulate the worst-case expectation $\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{h(x,\xi)<0\}}\right]$.

The next theorem combines Lemma 1 and Theorem 3 to provide the deterministic reformulation of the distributionally robust chance constraint (1b) in the continuum of realizations case.

THEOREM 4. *(**DRCCP** Reformulation for Continuum of Realizations) Suppose the uncertainty set Ω is closed and the set $\{h(x,\xi)<0\}$ has a non-empty intersection with the uncertainty set Ω. Let the indicator loss function defined by* (26)*. For any given $\theta \geq 0$, the distributionally robust chance constraint* (1b) *with continuum of realizations can be reformulated with the following set of inequalities $\mathcal{Z}^{CR}$:*

$$(\lambda,s,z,v)\in\mathcal{Z}^{CR} = \begin{cases} \lambda\theta + \dfrac{1}{N}\sum_{i=1}^{N} s_i \leq \epsilon & \text{(29a)} \\ \lambda \geq 0 & \text{(29b)} \\ s_i \geq 0, \quad \forall i\in\mathcal{N} & \text{(29c)} \\ \left[-l^{(1)}\right]^*(z_i - v_i) + \sigma_\Omega(v_i) - z_i^T\xi_i^0 \leq s_i, \forall i\in\mathcal{N} & \text{(29d)} \\ ||z_i||_* \leq \lambda, \quad \forall i\in\mathcal{N} & \text{(29e)} \end{cases}$$

Proof. Substituting (25a) for $\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{h(x,\xi)<0\}}\right]$ gives (29a). Non-negativity imposed on λ gives (29b). Recall $\mathcal{K} := \{1,2\}$, with $l^{(1)}(x,\xi) = 1-\chi_{\{h(x,\xi)<0\}}(\xi)$ and $l^{(2)}(x,\xi)=0$. The conjugate function $[-l^{(2)}]^*$ is $=0$, if $\xi=0$; $=\infty$, otherwise. The constraints involving $k=2$ can then be disregarded. We suppress the index k to ease the notation, and obtain (29d) and (29e) directly from (25b) and (25c). □

To further specify the reformulation, we need to determine the conjugate function $[-l^{(k)}]^*$ of the loss function $l(x,\xi)$ and the support function σ_Ω depending on the form of Ω defined as: $\Omega = \{\xi : C\xi \leq d\}$. Recall that ξ denotes a scalar in models with random right-hand side and is a M-dimensional vector in models with random technology vector. In a similar abuse of notation, $C\in\mathbb{R}_+$, $d\in\mathbb{R}$, and Ω is a halfspace in models with random right-hand side, while in models with random technology vector, $C\in\mathbb{R}_+^{r\times M}$ is a matrix, $d\in\mathbb{R}^r$ is a vector, and the support Ω is a polytope.

In the next subsections, we consider two functional forms for the stochastic inequalities defining the set $\{\xi : h(x,\xi)<0\}$ and define for each the specific formulation of the set $\mathcal{Z}^{CR}$ in Theorem (4). We first give a generic reformulation with bilinear terms. Next, we derive a mixed-integer conic programming reformulation (resp., outer approximation) for models with random right-hand side (resp., random technology vector). Finally, we define the conditions under which the outer approximation for the model with random technology matrix is equivalent to **DRCCP**.

4.1. Random Right-Hand Side

We study first the case with linear uncertainty in the right-hand side: $\{\xi : h(x,\xi) < 0\} = \{\xi : a^T x < \xi\}$, with $\xi \in \mathbb{R}$, $C \in \mathbb{R}_+$ and $d \in \mathbb{R}$. To ease the notation, we introduce a user-defined small positive number δ such that $a^T x + \delta \leq \xi$. Here too, we introduce binary variables in the proposed reformulation. Corollary 3 follows from Theorem 4.

COROLLARY 3. *(Linear Uncertainty in Right-Hand Side) The distributionally robust chance constraint*

$$\sup_{\mathbb{P}\in\mathcal{D}_W^{CR}} \mathbb{P}\left\{\xi : a^T x < \xi\right\} \leq \epsilon \tag{30}$$

can be reformulated with the set of constraints $\mathcal{Z}_R^{CR}$*:*

$$(x,\lambda,\beta,\gamma,s) \in \mathcal{Z}_R^{CR} = \begin{cases} 1 + \beta_i(a^T x + \delta - \xi_i^0) + \gamma_i(d - C\xi_i^0) \leq s_i, \forall i \in \mathcal{N} & \text{(31a)} \\ ||\beta_i + C\gamma_i||_* \leq \lambda, \quad \forall i \in \mathcal{N} & \text{(31b)} \\ \gamma_i \geq 0, \quad \forall i \in \mathcal{N} & \text{(31c)} \\ \beta_i \leq 0, \quad \forall i \in \mathcal{N} & \text{(31d)} \\ \text{(29a)} - \text{(29c)} \end{cases}$$

In the continuum of realizations case, the problem

$$\min_{x,\lambda,\beta,\gamma,s} \left\{g(x) : (x,\lambda,\beta,\gamma,s) \in \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N : (x,\lambda,\beta,\gamma) \in \mathcal{X} \cap \mathcal{Z}_R^{CR}\right\}$$

is equivalent to **DRCCP** *for linear uncertainty in the right-hand side.*

Proof. The uncertainty set is defined by $\Omega = \{\xi \in \mathbb{R} : C\xi \leq d\}$. We obtain the support function $\sigma_\Omega(v_i)$:

$$\sigma_\Omega(v_i) = \begin{cases} \sup\limits_{\xi} & v_i\xi_i \\ \text{s.t.} & C\xi_i \leq d \end{cases} = \begin{cases} \inf\limits_{\gamma\geq 0} & \gamma_i d \\ \text{s.t.} & C\gamma_i = v_i \end{cases} \tag{32}$$

where $v_i \in \mathbb{R}$, and $\gamma_i \in \mathbb{R}_+$ is the dual variable corresponding to $C\xi_i \leq d, \forall i \in \mathcal{N}$.

The set $\{\xi : h(x,\xi) < 0\}$ is given by $\{\xi : a^T x < \xi\}$. Introducing δ, the set becomes $\{\xi : a^T x + \delta \leq \xi\}$ The conjugate function $[-l^{(1)}]^*(z_i - v_i), \forall i \in \mathcal{N}$ is:

$$[-l^{(1)}]^*(z_i - v_i) = \begin{cases} \sup\limits_{\xi} & (z_i - v_i)\xi_i + 1 \\ \text{s.t.} & a^T x + \delta \leq \xi_i \end{cases} = \begin{cases} \inf\limits_{\beta_i \leq 0} & \beta_i(a^T x + \delta) + 1 \\ \text{s.t.} & \beta_i = z_i - v_i \end{cases} \tag{33}$$

where $\beta_i \in \mathbb{R}_-$ is the dual variable of constraint $\xi_i \geq a^T x + \delta, \forall i \in \mathcal{N}$. Thus we have $z_i = \beta_i + v_i = \beta_i + C\gamma_i$, where $z_i \in \mathbb{R}, \forall i \in \mathcal{N}$. Substituting (32), (33) and z_i into $\mathcal{Z}^{CR}$ gives the constraints defining $\mathcal{Z}_R^{CR}$. □

The above problem includes bilinear terms involving the decision variables x and β_i in (31a) and is non-convex. We propose an MISOCP reformulation in Theorem 5.

THEOREM 5. *(MISOCP Reformulation for DRCCP with Linear Uncertainty in Right-Hand Side) Suppose the uncertainty set Ω is defined by $\Omega = \{\xi : C\xi \leq d, C \geq 0\}$. Let L_x and U_x denote the lower and upper bound vectors of x. Let $\xi_i'^0 = \xi_i^0 - \delta, \forall i \in \mathcal{N}$. Define G_i as the positive constant:*

$$G_i = \xi_i'^0 - \sum_{m\in\mathcal{M}:a_m\geq 0} a_m L_{x_m} - \sum_{m\in\mathcal{M}:a_m<0} a_m U_{x_m}\ , \quad \forall i \in \mathcal{N}\ .$$

The distributionally robust chance constraint (30) *can be reformulated with the set of constraints $\mathcal{Z}_{RSO}^{CR}$:*

$$(x,\lambda,\beta,\gamma,s,y,w') \in \mathcal{Z}_{RSO}^{CR} = \begin{cases} 1 - w_i' \leq s_i, \quad \forall i \in \mathcal{N} & \text{(34a)} \\ w_i' \in \{0,1\}, \quad \forall i \in \mathcal{N} & \text{(34b)} \\ -\beta_i \leq \lambda, \quad \forall i \in \mathcal{N} & \text{(34c)} \\ \sum_{i\in\mathcal{N}} w_i' \geq \lceil N(1-\epsilon) \rceil & \text{(34d)} \\ y_i \geq 0, \quad \forall i \in \mathcal{N} & \text{(34e)} \\ a^T x - \xi_i'^0 \leq 2y_i, \quad \forall i \in \mathcal{N} & \text{(34f)} \\ a^T x - \xi_i'^0 + (1-w_i')G_i \geq 2y_i, \quad \forall i \in \mathcal{N} & \text{(34g)} \\ G_i w_i' \geq y_i, \quad \forall i \in \mathcal{N} & \text{(34h)} \\ -2\beta_i y_i \geq {w_i'}^2, \quad \forall i \in \mathcal{N} & \text{(34i)} \\ \text{(29a)} - \text{(29c)}, \text{(31c)} - \text{(31d)} \end{cases}$$

In the continuum of realizations case, the mixed-integer second-order cone programming problem

$$\min_{x,\lambda,\beta,\gamma,s,y,w'} \{g(x) : (x,\lambda,\beta,\gamma,s,y,w') \in \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \times \{0,1\}^N : (x,\lambda,\beta,\gamma,s,y,w') \in \mathcal{X} \cap \mathcal{Z}_{RSO}^{CR}\}$$

is equivalent to **DRCCP** *for linear uncertainty in the right-hand side.*

Proof. The set of constraints $\mathcal{Z}_{RSO}^{CR}$ is derived from $\mathcal{Z}_{R}^{CR}$. Since all training samples from the empirical distribution belong to the uncertainty set, which implies $d - C\xi_i^0 \geq 0, \forall i \in \mathcal{N}$, and $C \in \mathbb{R}_+$, it is optimal to set $\gamma_i = 0$ in (31a) and (31b), which gives:

$$1 + \beta_i(a^T x + \delta - \xi_i^0) \leq s_i, \qquad \forall i \in \mathcal{N} \tag{35a}$$

$$||\beta_i||_* \leq \lambda, \qquad \forall i \in \mathcal{N} \tag{35b}$$

Since the dual norm of the l_1-norm is the l_∞-norm (see Boyd and Vandenberghe (2004)), we obtain:

$$||\beta_i||_* = \max_{\beta_i} \{|\beta_i| : \forall i \in \mathcal{N}\} \tag{35c}$$

The variable β_i is non-positive (see (31d)) since it is the dual variable for constraint $\xi_i \geq a^T x + \delta$. Integrating (35c) and (31d) gives (34c).

Since θ is a non-negative constant and $\lambda \geq 0$ (29b), it follows that, for any given λ and θ, each term s_i must be as small as possible for (29a) to hold, and therefore $\beta_i(a^T x + \delta - \xi_i^0)$ must be minimal (but no smaller than -1 since $s_i \geq 0$). This leads us to distinguish two scenarios for each constraint (35a) and its term $\beta_i(a^T x + \delta - \xi_i^0)$ (with $\beta_i \leq 0$ due to (31d)). With the definition of $\xi_i'^0$, we rewrite $\beta_i(a^T x + \delta - \xi_i^0)$ as $\beta_i(a^T x - \xi_i'^0)$.

(i) if $a^T x - \xi_i'^0 \geq 0$, then it is optimal to set $\beta_i(a^T x - \xi_i'^0) \leq -1$, giving $s_i = 0$;

(ii) if $a^T x - \xi_i'^0 < 0$, then it is optimal to set $\beta_i(a^T x - \xi_i'^0) = 0$ by enforcing $\beta_i = 0$, allowing for $s_i = 1$, which is the smallest value that s_i can take under scenario (ii).

We now define a binary variable $w' \in \{0,1\}^N$ and a nonnegative variable $y \in \mathbb{R}_+^N$, and incorporate linear (34e)-(34h) and second-order cone (34i) constraints to force each variable w_i' and β_i to take their ideal value defined by:

$$w_i' = -\beta_i(a^T x - \xi_i'^0) = \begin{cases} 1, \text{ if } a^T x - \xi_i'^0 \geq 0, \text{ and} & \beta_i = -1/(a^T x - \xi_i'^0). \\ 0, \text{ if } a^T x - \xi_i'^0 < 0, \text{ and} & \beta_i = 0. \end{cases} \tag{36}$$

Each w_i' can thus be understood as a binary variable indicating if a given solution x satisfies or not the constraint $a^T x - \xi_i'^0 \geq 0$ under the atom ξ_i^0.

Consider the above two scenarios:

(i) If $a^T x - \xi_i'^0 > 0$, then $y_i \geq (a^T x - \xi_i'^0)/2 > 0$ due to (34f), which in turn implies $w_i' = 1$ due to (34h) and $y_i \leq (a^T x - \xi_i'^0)/2$ due to (34g). Therefore, $y_i = (a^T x - \xi_i'^0)/2$ and the convex quadratic constraint (34i) forces $\beta_i \leq -\frac{1}{a^T x - \xi_i'^0}$. It follows that (36) holds under (i).

(ii) If $a^T x - \xi_i'^0 < 0$, we have $w_i' = 0$ from (34g), $y_i = 0$ from (34h), and β_i can take value 0. It follows that (36) holds under (ii).

Constraint (34a) is obtained by substituting $\beta_i(a^T x - \xi_i'^0)$ by $-w_i'$ due to (36). The constraint (34d) holds if the aggregate probability of the satisfied atoms must be at least equal to $1 - \epsilon$, which is required for the distributionally robust chance constraint. Therefore, (34d) is valid and can be added to strengthen the formulation.

Note that we multiply y_i by 2 in (34f) and (34g) to obtain the canonical form of a rotated second constraint in (34i). □

Next, we propose an LP outer approximation of the distributionally robust chance constraint (30) obtained using the McCormick envelop relaxation of the bilinear term $\beta_i x$. Lemma 2 specifies the conditions under which the LP outer approximation is an equivalent reformulation.

THEOREM 6. *(LP Outer Approximation) Let* L_x *and* U_x *denote the lower and upper bound vectors of* x *and* L_β *be the lower bound of* β*. The distributionally robust chance constraint* (30) *can be relaxed by the set of constraints* $\mathcal{Z}^{CR}_{RLP}$*:*

$$(x,\lambda,\beta,\gamma,s,\phi)\in\mathcal{Z}^{CR}_{RLP}=\begin{cases}1-\beta_i\xi_i'^0+\sum_{m\in\mathcal{M}}a_m\phi_{mi}\le s_i, \quad \forall i\in\mathcal{N} & \text{(37a)}\\ \phi_{mi}\le U_{x_m}\beta_i+x_mL_{\beta_i}-U_{x_m}L_{\beta_i}, \\ \qquad\qquad \forall m\in\mathcal{M}, i\in\mathcal{N} & \text{(37b)}\\ \phi_{mi}\le L_{x_m}\beta_i, \quad \forall m\in\mathcal{M}, i\in\mathcal{N} & \text{(37c)}\\ \phi_{mi}\ge U_{x_m}\beta_i, \quad \forall m\in\mathcal{M}, i\in\mathcal{N} & \text{(37d)}\\ \phi_{mi}\ge L_{x_m}\beta_i+x_mL_{\beta_i}-L_{x_m}L_{\beta_i}, \\ \qquad\qquad \forall m\in\mathcal{M}, i\in\mathcal{N} & \text{(37e)}\\ \text{(29a)}-\text{(29c)}, \text{(31c)}-\text{(31d)}, \text{(34c)}\end{cases}$$

Proof. We employ a procedure similar to that used in Theorem 5 to tackle the dual norm in (31b) and obtain (34c). We next utilize the McCormick envelop to relax the bilinear term $\beta_i x_m$ in (31a). We introduce the auxiliary variables $\phi_{mi}:=x_m\beta_i, \forall i\in\mathcal{N}, m\in\mathcal{M}$, and (37a) is equivalent to (31a). Since $x\in[L_x,U_x]$ and $\beta\in[L_\beta,0]$ are bounded, we can derive the McCormick inequalities (37b)-(37e) to represent the bilinear term $x_m\beta_i$. □

We emphasize here that there are major differences between the formulations presented in Theorems 5 and 6. Theorem 5 presents a mixed-integer second-order cone programming scheme that equivalently reformulates the bilinear term $\beta_i(a^Tx-\xi_i'^0)$ and thus provides an exact reformulation to the distributionally robust chance constraint (30). Theorem 6 employs the McCormick approach to relax the bilinear term $\beta_i x_m$, which gives a linear programming relaxation (instead of an exact reformulation). The feasible set $\mathcal{Z}^{CR}_{RLP}$ represents the outer approximation of the chance constraint (39) due to the McCormick relaxation. The next lemma provides conditions under which the relaxation problem is exact and the feasible set $\mathcal{Z}^{CR}_{RLP}$ is equivalent to that of the chance constraint.

LEMMA 2. *(Exactness Conditions for LP Outer Approximation) The set* $\mathcal{Z}^{CR}_{CLP}$ *is equivalent to the feasible area defined by the distributionally robust constraint* (39) *if the decision variables* x *are bounded integer variables. The feasible region* $\mathcal{Z}^{CR}_{CLP}$ *is then represented by a set of mixed-integer linear programming (MILP) inequalities.*

If $x\in\{0,1\}^M$ is binary, the McCormick envelop relaxation provides an exact reformulation for each bilinear term $x\beta$. If $x\in\mathbb{Z}^M_+\cup[0,U_x]^M$ is a general integer variable, we can rewrite x as a sum of binary variables and apply McCormick's approach to each resulting term.

We explain in Theorem 7 how to set the lower bound for the the dual variable β.

THEOREM 7. *Let $\tau > 0$ be an infinitesimal non-negative number. The lower bound L_{β_i} of the dual variable β_i is:*

$$L_{\beta_i} = \begin{cases} \kappa^* & \textit{if } \exists x : a^T x - \xi_i^0 > 0 \\ 0, & \textit{otherwise} \end{cases}, \tag{38}$$

where κ^* is the optimal objective value of the linear programming problem:

$$\begin{aligned} \kappa^* = \min \quad & a^T x - \xi_i^0 > 0 \\ \textit{s.t.} \quad & a^T x - \xi_i^0 \geq 0 + \tau \\ & x \in \mathcal{X} \end{aligned}$$

Proof. Since $d - C\xi_i^0 \geq 0$, $\forall i \in \mathcal{N}$ and $C \in \mathbb{R}_+$, we know from Theorem 5 that it is optimal to set $\gamma_i = 0$ in (31a) and (31b), which gives (35a) and (35b). We distinguish two scenarios depending on the terms $\beta_i(a^T x - \xi_i^0)$ in (44a):
(i) If $a^T x - \xi_i^0 > 0$, it is optimal to set $\beta_i = -\frac{1}{a^T x - \xi_i^0}$ and the lower bound of β_i can then be obtained by solving any of the two equivalent problems

$$\min -\frac{1}{a^T x - \xi_i^0} : \; a^T x - \xi_i^0 > 0, \; x \in \mathcal{X} \; \Leftrightarrow \; \max \; \xi_i^0 - a^T x : \; a^T x - \xi_i^0 \geq 0 + \tau, \; x \in \mathcal{X}$$

with the right-side maximization problem taking the form of an LP problem.
(ii) If $a^T x - \xi_i^0 \leq 0$, it is optimal to set $\beta_i(a^T x - \xi_i^0) = 0$. It is further shown in the proof of Theorem 5 that $\beta_i = 0$ is optimal in this case. □

4.2. Random Technology Vector

We next present the reformulation for the case with linear uncertainty in the technology vector, in which the explicit form of the set $\{\xi : h(x, \xi) < 0\}$ is $\{\xi : \xi^T x < b\}$ with $\xi \in \mathbb{R}^M$, $C \in \mathbb{R}_+^{r \times M}$, and $d \in \mathbb{R}^r$. Let δ' be an infinitesimal positive number such that $\xi^T x \leq b - \delta'$.

COROLLARY 4. *(Linear Uncertainty in Random Technology Vector) The distributionally robust chance constraint*

$$\sup_{\mathbb{P} \in \mathcal{D}_W^{CR}} \mathbb{P}\left\{\xi : \xi^T x < b\right\} \leq \epsilon, \tag{39}$$

can be reformulated with the set of constraints $\mathcal{Z}_L^{CR}$:

$$(x, \lambda, \eta, \gamma, s) \in \mathcal{Z}_L^{CR} = \begin{cases} 1 + \eta_i(b - \delta' - x^T \xi_i^0) + \gamma_i^T (d - C\xi_i^0) \leq s_i, \forall i \in \mathcal{N} & \text{(40a)} \\ ||x\eta_i + C^T \gamma_i||_* \leq \lambda, \quad \forall i \in \mathcal{N} & \text{(40b)} \\ \gamma_i \geq 0, \quad \forall i \in \mathcal{N} & \text{(40c)} \\ \eta_i \geq 0, \quad \forall i \in \mathcal{N} & \text{(40d)} \\ \text{(29a)} - \text{(29b)} \end{cases}$$

In the continuum of realizations case, the problem

$$\min_{x,\lambda,\eta,\gamma,s} \left\{ g(x) : (x,\lambda,\eta,\gamma,s) \in \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N : (x,\lambda,\eta,\gamma,s) \in \mathcal{X} \cap \mathcal{Z}_L^{CR} \right\}$$

is equivalent to **DRCCP** *under linear uncertainty in the random technology vector.*

Proof. The support function $\sigma_\Omega(v_i)$ is derived as in (32) in Corollary 3. Since ξ is here a M-dimensional vector, the support function $\sigma_\Omega(v_i)$ reads:

$$\sigma_\Omega(v_i) = \begin{cases} \sup_{\xi} & v_i^T \xi_i \\ \text{s.t.} & C\xi_i \le d \end{cases} = \begin{cases} \inf_{\gamma \ge 0} & \gamma_i^T d \\ \text{s.t.} & C^T \gamma_i = v_i \end{cases} \tag{41}$$

where $\gamma_i \in \mathbb{R}_+^r$ is the dual variable vector corresponding to constraints $C\xi_i \le d, \forall i \in \mathcal{N}$.
The conjugate function $[-l^{(1)}]^*(z_i - v_i), \forall i \in \mathcal{N}$ is:

$$[-l^{(1)}]^*(z_i - v_i) = \begin{cases} \sup_{\xi} & (z_i - v_i)^T \xi_i + 1 \\ \text{s.t.} & x^T \xi_i \le b - \delta' \end{cases} = \begin{cases} \inf_{\eta \ge 0} & (b - \delta')\eta_i + 1 \\ \text{s.t.} & x\eta_i = z_i - v_i \end{cases} \tag{42}$$

where $\eta_i \in \mathbb{R}$ is the dual variable for constraint $x^T \xi_i \le b - \delta'$, $\forall i \in \mathcal{N}$. Thus we have $z_i = x\eta_i + C^T \gamma_i$, where $x, z_i \in \mathbb{R}^M$. Substituting (41), (42) and z_i into $\mathcal{Z}^{CR}$ gives the set of constraints defining $\mathcal{Z}_L^{CR}$. □

As for the case with linear uncertainty in the right-hand side, problem in Corollary 4 includes bilinear terms involving x and η_i in constraint (40a) and is non-convex. The complexity of $\mathcal{Z}_L^{CR}$ is further compounded by the presence of additional bilinear products $x\eta_i$ appearing in the dual norm term in (40b). Theorem 8 provides a mixed-integer second-order cone relaxation of the distributionally robust chance constraint with random technology matrix and continuum of realizations uncertainty type.

THEOREM 8. *(MISOCP Outer Approximation) Let L_x, U_x denote lower and upper bound vectors of x and U_η be the upper bound of η. Let $b' = b - \delta'$. Define $G_i' > 0, \forall i \in \mathcal{N}$ as the positive constant:*

$$G_i' = b' - \sum_{m \in \mathcal{M}: \xi_{im}^0 \ge 0} \xi_{im}^0 L_{x_m} - \sum_{m \in \mathcal{M}: \xi_{im}^0 < 0} \xi_{im}^0 U_{x_m}, \quad \forall i \in \mathcal{N}.$$

The distributionally robust chance constraint (39) *with continuum of realizations can be relaxed by the set of constraints* $\mathcal{Z}_{CSO}^{CR}$*:*

$$
(x,\lambda,\beta,\gamma,s,y,w',\phi) \in \mathcal{Z}_{CSO}^{CR} = \begin{cases}
1 - w'_i \leq s_i, \quad \forall i \in \mathcal{N} & \text{(43a)} \\
w'_i \in \{0,1\}, \quad \forall i \in \mathcal{N} & \text{(43b)} \\
\sum_{i \in \mathcal{N}} w'_i \geq \lceil (1-\epsilon)N \rceil \,, & \text{(43c)} \\
y_i \geq 0, \quad \forall i \in \mathcal{N} & \text{(43d)} \\
b' - x^T \xi_i^0 + 2y_i \geq 0, \quad \forall i \in \mathcal{N} & \text{(43e)} \\
b' - x^T \xi_i^0 + 2y_i \leq (1-w'_i)G'_i, \ \forall i \in \mathcal{N} & \text{(43f)} \\
G'_i w'_i \geq y_i, \quad \forall i \in \mathcal{N} & \text{(43g)} \\
2\eta_i y_i \geq {w'_i}^2, \quad \forall i \in \mathcal{N} & \text{(43h)} \\
\phi_{mi} \leq \lambda, \quad \forall m \in \mathcal{M}, i \in \mathcal{N} & \text{(43i)} \\
\phi_{mi} \leq U_{x_m} \eta_i, \quad \forall m \in \mathcal{M}, i \in \mathcal{N} & \text{(43j)} \\
\phi_{mi} \leq L_{x_m} \eta_i + U_\eta x_m - L_{x_m} U_\eta, & \\
\qquad\qquad \forall m \in \mathcal{M}, i \in \mathcal{N} & \text{(43k)} \\
\phi_{mi} \geq L_{x_m} \eta_i, \quad \forall m \in \mathcal{M}, i \in \mathcal{N} & \text{(43l)} \\
\phi_{mi} \geq U_\eta x_m + U_{x_m} \eta_i - U_{x_m} U_\eta, & \\
\qquad\qquad \forall m \in \mathcal{M}, i \in \mathcal{N} & \text{(43m)} \\
\text{(29a)} - \text{(29c)}, \text{(40c)} - \text{(40d)} &
\end{cases}
$$

Proof. Since $d - C\xi_i^0 \geq 0$, $\forall i \in \mathcal{N}$, given that the scenarios of the empirical distribution all belong to the uncertainty set and $C \in \mathbb{R}_+^{r \times M}$, it is optimal to set $\gamma_i = 0$ in (40a) and (40b), which can hence be rewritten as:

$$
1 + \eta_i(b - \delta' - x^T \xi_i^0) \leq s_i, \qquad \forall i \in \mathcal{N} \tag{44a}
$$

$$
||x\eta_i||_* \leq \lambda, \qquad \forall i \in \mathcal{N} \tag{44b}
$$

We decompose the remaining part of the proof into two steps. We first derive the mixed-integer second-order cone programming formulations to represent constraints (44a) before linearizing the bilinear terms in (44b) using the McCormick inequalities McCormick (1976).

Step 1. Mixed-integer second-order cone representation. Recall that $b' = b - \delta'$. Hence, the term $\eta_i(b - \delta' - x^T \xi_i^0)$ can be rewritten as $\eta_i(b' - x^T \xi_i^0)$. Let's focus on the terms $\eta_i(b' - x^T \xi_i^0)$ in (44a) and distinguish two scenarios:

(i) if $b' - x^T \xi_i^0 \leq 0$, then it is optimal to set $\eta_i \geq \frac{1}{x^T \xi_i^0 - b'}$, which allows for $s_i = 0$.

(ii) if $b' - x^T\xi_i^0 > 0$, it is optimal to set $\eta_i(b' - x^T\xi_i^0) = 0$ with $\eta_i = 0$, allowing for $s_i = 1$, i.e., the smallest possible value for s_i given the non-negativity of η_i and $(b' - x^T\xi_i^0)$ under scenario (ii).

We now introduce a vector of binary $w' \in \{0,1\}^N$ and nonnegative $y \in \mathbb{R}_+^N$ variables. The binary variables are used to replace the term $-\eta_i(b' - x^T\xi_i^0)$. Based on (i) and (ii), we incorporate a set of second-order cone constraints that force each binary variable w_i' to take the optimal values presented below, namely:

$$w_i' = -\eta_i(b' - x^T\xi_i^0) = \begin{cases} 1, & \text{if } b' - x^T\xi_i^0 \leq 0, \text{ and } \eta_i = -1/(b' - x^T\xi_i^0). \\ 0, & \text{if } b' - x^T\xi_i^0 > 0, \text{ and } \eta_i = 0. \end{cases} \tag{45}$$

Each binary variable w_i' can be interpreted as indicative of whether or not a given solution x satisfies the constraint $b' - x^T\xi_i^0 < 0$ under the atom ξ_i^0:

(i) If $b' - x^T\xi_i^0 \geq 0$, (43f) forces $w_i' = 0$ which, in turn, leads to $y_i = 0$ due to (43g).

(ii) If $b' - x^T\xi_i^0 < 0$, we have $y_i \geq (x^T\xi_i^0 - b')/2 > 0$ from (43e) and $w_i' = 1$ from (43g). Therefore, if $b' - x^T\xi_i^0 < 0 \Leftrightarrow w_i' = 1$, the nonlinear inequality

$$2\eta_i y_i \geq w_i', \quad \forall i \in \mathcal{N} \tag{46}$$

forces $\eta_i \geq \frac{1}{x^T\xi_i^0 - b'}$. On the other hand, if $b' - x^T\xi_i^0 > 0 \Leftrightarrow w_i' = 0$, η_i can take value 0.

Since $w_i' \in \{0,1\}$, then $w_i' = {w_i'}^2$, we can replace the right-hand side of (46) by ${w_i'}^2$, which gives the canonical rotated second order cone constraint (43h).

Furthermore, (44a) is simplified as (43a) by substituting w_i' based on (45). To see the equivalence, consider the following:

i) If $b' - x^T\xi_i^0 \geq 0$, $w_i' = 0$ and the optimal value of η_i is 0. Thus, (43a) becomes: $1 = 1 - w_i' \leq s_i$.

ii) If $b' - x^T\xi_i^0 < 0$, we have $w_i' = 1$, $y_i \geq (x^T\xi_i^0 - b')/2$, $\eta_i \geq \frac{1}{x^T\xi_i^0 - b'}$, and $\eta_i(b' - x^T\xi_i^0)$ in the left side of (43a) is then smaller than -1, which allows its replacement by $-w_i'$.

Owing to the definition of w_i' (45) which takes value 1 when $b' - x^T\xi_i^0 < 0$ and 0 otherwise and to the requirement that the aggregate probability of the satisfied atoms must be at least equal to $1 - \epsilon$, (43c) is valid and can be added to strengthen the formulation.

Step 2. McCormick envelop relaxation. The dual of the l_1 norm is the l_∞ norm; thus, (44b) becomes

$$||x\eta_i||_* = \max_{x,\eta} \{|x_m\eta_i| : \forall m \in \mathcal{M}\} \leq \lambda, \quad \forall i \in \mathcal{N} \tag{47}$$

The dual variable η corresponding to $x^T\xi \leq b'$ in the derivation of the conjugate function is nonnegative and upper-bounded by U_η. Due to (45), η_i either takes value 0 if $b - x^T\xi_i^0 \leq 0$ or $-1/(b' - x^T\xi_i^0)$ if $b' - x^T\xi_i^0 > 0$. To derive the envelope of the bilinear terms $x_m\eta_i, m \in \mathcal{M}, i \in \mathcal{N}$, we introduce a set of auxiliary variables $\phi_{mi} := x_m\eta_i$ constrained by the McCormick inequalities and obtain the convex relaxation of the bilinear term $x_m\eta_i$ represented by (43j)-(43m). Substituting $\phi_{mi} := x_m\eta_i$ into (47) gives (43i). □

THEOREM 9. *(LP Outer Approximation) Let L_x, U_x denote lower and upper bound vectors of x and U_η be the upper bound of η. The distributionally robust chance constraint* (39) *with continuum of realizations can be relaxed by the set of constraints $\mathcal{Z}^{CR}_{CLP}$:*

$$(x,\lambda,\beta,\gamma,s,\phi) \in \mathcal{Z}^{CR}_{CLP} = \begin{cases} 1+\eta_i b' - \sum_{m\in\mathcal{M}} \xi^0_{mi}\phi_{mi} \le s_i, \quad \forall i \in \mathcal{N} \\ \text{(29a)} - \text{(29c)}, \text{(40c)} - \text{(40d)}, \text{(43i)} - \text{(43m)} \end{cases} \tag{48a}$$

Proof. Similar to Theorem 5, $\mathcal{Z}^{CR}_{CLP}$ is obtained by using the McCormick envelop relaxation of the bilinear term $\eta_i x_m$ and by introducing a set of auxiliary variables $\phi_{mi} := x_m\beta_i$ defined by (43j)-(43m). Substituting $\phi_{mi} = x_m\eta_i$ into (40a) gives (48a). □

We note that either the feasible set $\mathcal{Z}^{CR}_{CSO}$ or $\mathcal{Z}^{CR}_{CLP}$ outer-approximates the feasible area of the chance constraint (39) due to the McCormick relaxation of bilinear terms $x\eta_i$. We next provide some conditions under which the above relaxation problem is exact and the feasible set defined by $\mathcal{Z}^{CR}_{CSO}$ or $\mathcal{Z}^{CR}_{CLP}$ is equivalent to that of the distributionally robust chance constraint.

LEMMA 3. *(Exactness Conditions for MISOCP or LP Outer Approximation) The set $\mathcal{Z}^{CR}_{CSO}$ or $\mathcal{Z}^{CR}_{CLP}$ is equivalent to the feasible area defined by the distributionally robust constraint* (39) *if the decision variables x are bounded integer variables.*

If $x \in \{0,1\}^M$ is binary, the McCormick envelop relaxation provides an exact reformulation for each bilinear term $x\eta$. If $x \in \mathbb{Z}^M_+ \cup [0,U_x]^M$ is a general integer variable, we rewrite it as a sum of binary variables and apply McCormick's approach to each resulting term.

We explain in Theorem 10 how to set the upper bound for the the dual variable η.

THEOREM 10. *Let $\tau' \ge 0$ be a user-defined small non-negative number. The upper bound U_{η_i} of the dual variable η_i is:*

$$U_{\eta_i} = \begin{cases} \kappa'^* & \textit{if } \exists x : x^T\xi^0_i - b > 0 \\ 0, & \textit{otherwise} \end{cases}, \tag{49}$$

where κ'^* is the optimal objective value of the linear programming problem:

$$\begin{aligned} \kappa'^* = \min \quad & x^T\xi^0_i - b \\ \textit{s.t.} \quad & b - x^T\xi^0_i \le 0 - \tau' \\ & x \in \mathcal{X} \end{aligned}$$

Proof. Since $d - C\xi^0_i \ge 0, \forall i \in \mathcal{N}$ and $C \in \mathbb{R}^{r\times M}_+$, we know from Theorem 8 that it is optimal to set $\gamma_i = 0$ in (40a) and (40b), which gives (44a) and (44b). We distinguish two scenarios depending on the terms $\eta_i(b - x^T\xi^0_i)$ in (44a):

(i) If $b - x^T\xi_i^0 < 0$, it is optimal to set $\eta_i = \frac{1}{x^T\xi_i^0 - b}$ and the upper bound of η_i can then be obtained by solving any of the two equivalent problems

$$\max \frac{1}{x^T\xi_i^0 - b}: \; b - x^T\xi_i^0 < 0, \; x \in \mathcal{X} \quad \Leftrightarrow \quad \min \; x^T\xi_i^0 - b: \; b - x^T\xi_i^0 \leq 0 - \tau', \; x \in \mathcal{X}$$

with the right-side minimization problem taking the form of a LP problem.
(ii) If $b - x^T\xi_i^0 \geq 0$, it is optimal to set $\eta_i(b - x^T\xi_i^0) = 0$. It is further shown in the proof of Theorem 8 that $\eta_i = 0$ is optimal in this case. □

5. Extension to Distributionally Robust Joint Chance Constraints

In this section, we extend our proposed framework to distributionally robust optimization problems with joint chance constraints. The generic formulation of the distributionally robust joint chance-constrained problem **DRJCCP** is:

$$\begin{aligned}
\textbf{DRJCCP:} \quad \min_x \quad & g(x) \\
\text{s.t.} \quad & \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\xi : h_t(x,\xi) \geq 0, \; \forall t \in \mathcal{T}\} \geq 1 - \epsilon, \\
& x \in \mathcal{X}
\end{aligned} \tag{50}$$

The **DRJCCP** model seeks a solution $x \in \mathcal{X}$ that minimizes a convex cost function $g(x)$ while ensuring, via the ambiguous joint chance constraint (50), that the worst-case probability of the joint event $\{h_t(x,\xi) \geq 0, \forall t \in \mathcal{T}\}$ exceeds the predefined probability threshold $1 - \epsilon$ within the ambiguity set $\mathcal{D}$. Constraint (50) represents the joint chance constraint when $T \geq 2$ and reduces to a single chance constraint if $T = 1$.
Utilizing the complement rule, the joint chance constraint (50) is equivalent to

$$\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\xi : h_t(x,\xi) < 0, \; \exists t \in \mathcal{T}\} \leq \epsilon, \tag{51}$$

Owing to the equivalence between a chance constraint and an expectation one with indicator function, we rewrite (51) as:

$$\sup_{\mathbb{P} \in \mathcal{D}} \; \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{h_t(x,\xi) < 0, \; \exists t \in \mathcal{T}\}}\right] \leq \epsilon \tag{52}$$

where $\mathbb{1}_{\{h_t(x,\xi) < 0, \; \exists t \in \mathcal{T}\}}$ denotes the indicator function taking value 1 if $h_t(x,\xi) < 0, \exists t \in \mathcal{T}$ and 0 if $h_t(x,\xi) \geq 0, \forall t \in \mathcal{T}$.

5.1. DRJCCP – Case with Uncertain Probabilities

Analogous to Theorem 1, we give the reformulation for joint chance constraint under the uncertain probabilities case in Theorem 11.

THEOREM 11. *(DRJCCP Reformulation for Uncertain Probabilities) Let* $w_{jt} = \mathbb{1}_{\{h_t(x,\xi_j)<0\}}$, $\forall j \in \mathcal{N}, t \in \mathcal{T}$. *For any* $\theta \geq 0$, *the distributionally robust chance constraint* (51) *with uncertain probabilities can be reformulated with the set of mixed-integer linear inequalities* $\mathcal{Z}^{JUP}$

$$(x,\lambda,v,w) \in \mathcal{Z}^{JUP} = \begin{cases} \lambda c_{ij} + v_i \geq w_{jt}, \quad \forall i,j \in \mathcal{N} : i \neq j, t \in \mathcal{T} & \text{(53a)} \\ w_{jt} = \mathbb{1}_{\{h_t(x,\xi_j)<0\}}, \quad \forall j \in \mathcal{N}, t \in \mathcal{T} & \text{(53b)} \\ \text{(8a)} - \text{(8b)} \end{cases}$$

Proof. The value of the worst-case expectation can be obtained from:

$$\sup_{\mathbb{P}\in\mathcal{D}_W^{UP}} \mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{h_t(x,\xi)<0,\ \exists t\in\mathcal{T}\}}] = \sup_{\pi,p\geq 0} \sum_{j\in\mathcal{N}} p_j \mathbb{1}_{\{h_t(x,\xi_j)\leq 0,\ \exists t\in\mathcal{T}\}} \quad \text{s.t.} \quad \text{(9b)} - \text{(9d)} \tag{54}$$

Employing similar procedure in the proof of Theorem 1, we use Lagrangian duality to reformulate the above problem as:

$$\sup_{\mathbb{P}\in\mathcal{D}_W^{UP}} \mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{h_t(x,\xi)\leq 0,\ \exists t\in\mathcal{T}\}}] = \sup_{\pi\geq 0} \inf_{\lambda\geq 0,v} \sum_{i,j\in\mathcal{N}} \pi_{ij} \mathbb{1}_{\{h_t(x,\xi_j)\leq 0,\ \exists t\in\mathcal{T}\}} + \lambda\left(\theta - \sum_{i,j\in\mathcal{N}} \pi_{ij}c_{ij}\right) + \sum_{i\in\mathcal{N}} v_i \left(p_i^0 - \sum_{j\in\mathcal{N}} \pi_{ij}\right) \tag{55a}$$

$$= \inf_{\lambda\geq 0,v} \lambda\theta + \sum_{i\in\mathcal{N}} p_i^0 v_i + \begin{cases} 0, & \text{if } \lambda c_{ij} + v_i - \mathbb{1}_{\{h_t(x,\xi_j)\leq 0,\ \exists t\in\mathcal{T}\}} \geq 0, \forall i,j \in \mathcal{N} \\ +\infty, & \text{otherwise} \end{cases}. \tag{55b}$$

Note that

$$\mathbb{1}_{\{h_t(x,\xi_j)<0,\ \exists t\in\mathcal{T}\}} = \max_{t\in\mathcal{T}} \mathbb{1}_{\{h_t(x,\xi_j)<0\}} \tag{55c}$$

We thus can rewrite the constraint

$$\lambda c_{ij} + v_i - \mathbb{1}_{\{h_t(x,\xi_j)<0,\ \exists t\in\mathcal{T}\}} \geq 0, \quad \forall i,j \in \mathcal{N} : i \neq j \tag{55d}$$

equivalently as

$$\lambda c_{ij} + v_i \geq \max_{t\in\mathcal{T}} \mathbb{1}_{\{h_t(x,\xi_j)<0\}}, \quad \forall i,j \in \mathcal{N} : i \neq j \tag{55e}$$

We define a new set of binary variables w_{jt} such that

$$w_{jt} = \mathbb{1}_{\{h_t(x,\xi_j)<0\}} = \begin{cases} 1, & \text{if } h_t(x,\xi_j) < 0 \\ 0, & \text{if } h_t(x,\xi_j) \geq 0 \end{cases}. \tag{55f}$$

By substituting (55f) into the constraint (55e), we obtain the following equivalent constraint

$$\lambda c_{ij} + v_i \geq w_{jt}, \quad \forall i,j \in \mathcal{N} : i \neq j, \forall t \in \mathcal{T} \tag{55g}$$

Using (55b) to replace the worst-case expectation $\sup_{\mathbb{P}\in\mathcal{D}_W^{UP}} \mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{h(x,\xi)\leq 0\}}]$ into (52) gives (53a). Introducing the linkage between w_j and $\mathbb{1}_{\{h(x,\xi_j)\leq 0\}}$ leads to (53b). □

The key difference between the reformulations of joint and individual chance constraints is due to (53b), which stipulates that NT equality constraints must hold instead of N in the individual chance constraint case. The reformulation technique for individual chance constraints presented in Section 3 can otherwise be directly generalized to the joint case with appropriate notational modifications.

5.2. DRJCCP - Case with Continuum of Realizations

We now extend the reformulation framework presented in Section 4 for individual chance constraints under continuum of realizations case to the case of joint chance constraint. In order to do so, we introduce Lemma 4 that represents the indicator function corresponding to the joint chance constraint as a pointwise maximom.

LEMMA 4. *The indicator function $\mathbb{1}_{\{h_t(x,\xi)<0,\exists t\in\mathcal{T}\}}$ can be rewritten as the pointwise maximum of a finite number of concave functions:*

$$\mathbb{1}_{\{h_t(x,\xi_i)<0,\exists t\in\mathcal{T}\}} = \max_{t=1,\ldots,T+1} \{\underbrace{1-\chi_{\{h_t(x,\xi_i)<0\}}}_{l_i^{(t)}}, \underbrace{0}_{l_i^{(T+1)}}\} \tag{56a}$$

in which

$$\chi_{\{h_t(x,\xi_i)<0\}}(\xi) = \begin{cases} 0, \textit{ if } \ h_t(x,\xi_i)<0 \\ \infty, \textit{ otherwise} \end{cases} \tag{56b}$$

is the characteristic function of the (open) convex set defined by $h(x,\xi_i)<0$.

Proof. We note that

$$\mathbb{1}_{\{h_t(x,\xi_j)<0,\ \exists t\in\mathcal{T}\}} = \max_{t\in\mathcal{T}} \mathbb{1}_{\{h_t(x,\xi_j)<0\}} \tag{57a}$$

With the definition of characteristic function (56b), we have

$$\mathbb{1}_{\{h_t(x,\xi_i)<0\}} = \max\{1-\chi_{\{h_t(x,\xi_i)<0\}}(\xi), 0\} = \begin{cases} 1, \text{ if } \ h_t(x,\xi)<0 \\ 0, \text{ otherwise} \end{cases} \tag{57b}$$

We thus can rewrite (57a) as follows:

$$\mathbb{1}_{\{h_t(x,\xi_i)<0,\exists t\in\mathcal{T}\}} = \begin{cases} 1, \text{ if } \ h_t(x,\xi)<0, \exists t\in\mathcal{T} \\ 0, \text{ if } \ h_t(x,\xi)\geq 0, \forall t\in\mathcal{T} \end{cases} \tag{57c}$$

$$= \max_{t=1,\ldots,T+1} \{\underbrace{1-\chi_{\{h_t(x,\xi_i)<0\}}}_{l_i^{(t)}}, \underbrace{0}_{l_i^{(T+1)}}\} \tag{57d}$$

Since the characteristic function $\chi_{\{h_t(x,\xi_i)<0\}}$ of a convex set is convex (Boyd and Vandenberghe 2004), $l_i^{(t)}, \forall t \in \mathcal{T}$ and $l_i^{(T+1)}$ are concave, and $l(x,\xi_i) = \max_{t=1,\ldots,T+1} \{\underbrace{1-\chi_{\{h_t(x,\xi_i)<0\}}}_{l_i^{(t)}}, \underbrace{0}_{l_i^{(T+1)}}\}$ is the finite maxima of concave functions. □

The next theorem generalizes the reformulation proposed in Theorem 4 for individual chance constraints to the joint chance constraint setting. case.

THEOREM 12. *(**DRCCP** Reformulation for Continuum of Realizations) Suppose the uncertainty set Ω is closed and the set $\{h(x,\xi)<0\}$ has a non-empty intersection with the uncertainty set Ω. Let the indicator loss function defined by* (56a)*. For any given $\theta \geq 0$, the distributionally robust chance constraint* (52) *with continuum of realizations can be reformulated with the following set of inequalities $\mathcal{Z}^{JCR}$:*

$$(\lambda,s,z,v) \in \mathcal{Z}^{JCR} = \begin{cases} \left[-l^{(t)}\right]^* (z_{it}-v_{it}) + \sigma_\Omega(v_{it}) - z_{it}^T \xi_i^0 \leq s_i, \forall i \in \mathcal{N}, t \in \mathcal{T} & \text{(58a)} \\ ||z_{it}||_* \leq \lambda \ , \ \forall i \in \mathcal{N}, t \in \mathcal{T} & \text{(58b)} \\ \text{(29a)} - \text{(29c)}; \end{cases}$$

Under the continuum of realizations case, the difference between the reformulation for joint chance constraints and that for individual ones stems from the representation of the indicator function, which, in turn, impacts significantly the dimension of the constraint set. For individual chance constraints, the indicator function is defined as the maximum of two concave functions (see (26)), while, for joint chance constraints, the indicator function is defined as the pointwise maximum of $T+1$ concave functions (see (56a)). The reformulation proposed in Theorem 4 for individual chance constraints can thus be directly generalized for joint ones (see Theorem 12). The key difference is due to constraints (58a) and (58b) which require that NT inequality constraints must hold for the joint chance constraint to hold true (instead of N for individual chance constraints).

6. Computational Tests

6.1. Experimental Design

In this section, we conduct a series of numerical experiments to evaluate the computational efficiency of the proposed solution framework (Section 6.2) and show how to choose a proper value for the Wasserstein radius θ via cross-validation (Section 6.3). In doing so, we study a distributionally robust multidimensional knapsack problem (DRMKP) (see Cheng et al. (2014)) with joint chance constraints and in the continuum of realizations uncertainty case. When the decision variables are continuous, the proposed MISOCP (Theorem 8) and LP (Theorem 9) formulations are outer approximations of the chance constraint. When the decision variables are binary, the proposed MISOCP and LP (who becomes MILP) formulations are equivalent reformulations of the distributionally robust chance constraint. The following notations are adopted for the **DRMKP** problem. We consider T knapsacks, M items and $g \in \mathbb{R}_+^M$ is the vector of item values. Let

$\xi \in \mathbb{R}^M$ denote the vector of random item weights, and $b > 0$ be the capacity of the knapsack. The decision variable x_m represents the proportion of mth item to be picked. We let $x \in \mathcal{X} := [0,1]$ for the continuous **DRMKP**, while let $x \in \mathcal{X} := \{0,1\}$ for the binary **DRMKP**. The solution of continuous **DRMKP** can be regarded as outer approximations of the binary **DRMKP**. Using the above notational setting, the **DRMKP** problem reads:

$$\textbf{DRMKP:} \quad \max_{x} \quad g^* = g^T x \tag{59a}$$

$$\text{s.t.} \quad \inf_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\left\{\xi : \xi_t^T x \le b_t, \forall t \in \mathcal{T}\right\} \ge 1-\epsilon \tag{59b}$$

$$x \in \mathcal{X} \tag{59c}$$

To fit the distributionally robust chance constraint (59b) into our modeling and solution framework, we first reformulate (59b) as:

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\left\{\xi : \xi_t^T x > b_t, \exists t \in \mathcal{T}\right\} \le \epsilon \ ,$$

which, defining $\tilde{\xi} = -\xi$ and $\tilde{b} = -b$, is equivalent to:

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{P}\left\{\tilde{\xi} : \tilde{\xi}_t^T x < \tilde{b}_t, \exists t \in \mathcal{T}\right\} \le \epsilon \tag{60}$$

We shall utilize Theorems 8 and 9 to reformulate the above chance constraint (60).

The problem instances are created as follows. The random weights ξ are generated from a uniform distribution $\xi \sim U(5,15)$ and the item values are sampled from another uniform distribution $g \sim U(0,20)$. We consider a number of knapsacks $T = \{1,5\}$ (in which the chance constraint reduces to an individual one when $T = 1$), a number of items $M = \{10,20\}$ with data sample size $N = \{100,1000\}$, a risk tolerance level $\epsilon \in \{0.05, 0.10\}$, and a Wasserstein radius $\theta \in \{0.01, 0.03\}$. We define the knapsack capacity as $b = 10 \times M$. We create a total of 32 problem instances by considering the above combinations of values for the parameters. For the continuous **DRMKP**, we solve the MISOCP and LP outer approximations for each instance; for the binary **DRMKP**, we solve the MISOCP and MILP exact reformulations.

We use AMPL to formulate all problems. Each problem instance is solved with Gurobi 9.0 solver on a 64-bit laptop with Intel(R) Core(TM) i7-6600 CPU processor at 2.60GHz with 16GB RAM. We set the time limit to 3600 seconds. For each instance, we report the solution time and objective value obtained with each model.

6.2. Numerical Demonstration of Computational Efficiency

In this subsection, we assess the computational efficiency of the exact MISOCP and MILP reformulation to solve the binary **DRMKP** problems and of the relaxed MISOCP and LP approximations to solve the continuous **DRMKP** problems. For the binary **DRMKP** problem, the MISOCP and MILP reformulations are equivalent to the DRO problem and give the same solution (if solved within the time limit). Note also that for any given problem instance, the optimal solution of the continuous **DRMKP** problem can be regarded as an outer approximation of the corresponding binary **DRMKP** problem. The optimal value of the MISOCP and LP relaxations of the continuous problem provides therefore an upper bound (since these are maximization problems) on the optimal value of the binary problem. We will also assess in this subsection the tightness of the MISOCP and LP relaxations.

We denote by g^* the optimal objective value of the binary **DRMKP** problem that can be obtained by solving one of the exact (i.e., MISOCP or MILP) reformulation. Let g^*_{SO} and g^*_{LP} denote the optimal value of the continuous **DRMKP** problem obtained by solving the MISOCP and LP relaxation of the continuous **DRMKP** problem, respectively. Let GAP_{SO} (resp. GAP_{LP}) represent the optimality gap between the solution of the MISOCP (resp. LP) relaxation model (continuous **DRMKP**) and the solution of the ("true model") binary **DRMKP** problem, which is computed as:

$$GAP_{SO} = \frac{g^*_{SO} - g^*}{g^*} \ , \quad GAP_{LP} = \frac{g^*_{LP} - g^*}{g^*}$$

The numerical results for the **DRMKP** problem instances with single ($T = 1$) and joint ($T = 5$) chance constraints are displayed in Table 1 and 2, respectively. In these tables, we use “MISOCP”, “MILP”, “MISOCP-Approx” and “LP-Approx” to denote the exact MISOCP and MILP reformulations to solve binary **DRMKP** problems and the outer approximation MISOCP and LP formulations to solve continuous **DRMKP** problems, respectively. We use “Time” to denote the computational solution time in seconds. If the binary **DRMKP** is not solved to optimality within the time limit, the optimal objective value g^* and the corresponding gap values GAP_{SO} and GAP_{LP} are denoted as “–”. The row called ”Average” reports the average solution time and average GAP values with respect to the number of items M and number of data points N in the considered instances. The term “Overall Average” denotes the average value of solution time and gaps across all problem instances.

6.2.1. Results for Individual Chance Constraints Table 1 presents the results for **DRMKP** with individual chance constraint ($T = 1$) and show that all binary **DRMKP** models can be solved to optimality within 17 minutes (850 seconds). The MISOCP reformulation of the instance with 20 items and 1000 data points takes the longest time to solve, i.e., 831.86 seconds. For the exact reformulation methods, the MILP reformulation is for each instance solved faster than the MISOCP reformulation. On average, the solution of the MILP reformulation takes 43% of the solution time needed to solve the MISOCP reformulation (70.83s vs. 163.90s). A similar pattern prevails for the solution of the MISOCP and LP outer approximation problems of the continuous **DRMKP** problems. The LP-Approx relaxation dominates the MISOCP-Approx one in terms of solution time, as LP-Approx is solved, for each instance, much faster than MISOCP-Approx. On average, the LP-Approx solution time amounts to no more than 4.8% of the

Table 1 Numerical results with exact reformulation and outer approximation for binary and continuous DRMKP problems: Individual ($T=1$) chance constraint case

Parameters			Binary DRMKP			Continuous DRMKP					
				MISOCP	MILP	MISOCP-Approx			LP-Approx		
$T=1$	M	N	g^*	Time	Time	g^*_{SO}	GAP_{SO}	Time	g^*_{LP}	GAP_{LP}	Time
$\epsilon=0.05, \theta=0.01$	10	100	100	1.25	0.94	103.439	3.44%	0.77	107.994	7.99%	0.14
	10	1000	100	56.99	19.73	102.403	2.40%	32.16	107.994	7.99%	3.81
	20	100	220	4.64	2.67	221.304	0.59%	2.46	227.997	3.64%	0.39
	20	1000	215	543.37	388.12	215.853	0.40%	400.45	227.997	6.05%	13.09
$\epsilon=0.05, \theta=0.03$	10	100	100	1.78	1.01	102.952	2.95%	1.45	107.994	7.99%	0.06
	10	1000	100	48.81	14.63	101.777	1.78%	30.46	107.994	7.99%	2.92
	20	100	220	3.86	2.53	220.876	0.40%	2.77	227.998	3.64%	0.44
	20	1000	215	831.86	451.67	215.853	0.40%	720.65	227.997	6.05%	13.59
$\epsilon=0.10, \theta=0.01$	10	100	100	1.21	0.92	104.809	4.81%	0.73	107.995	8.00%	0.12
	10	1000	100	39.34	20.55	103.396	3.40%	28.13	107.995	8.00%	4.87
	20	100	220	2.65	1.76	222.546	1.16%	1.86	227.998	3.64%	0.20
	20	1000	220	423.93	113.57	221.217	0.55%	319.65	227.997	3.64%	17.47
$\epsilon=0.10, \theta=0.03$	10	100	100	1.58	0.94	104.407	4.41%	1.18	107.995	8.00%	0.08
	10	1000	100	53.40	26.44	103.349	3.35%	42.12	107.995	8.00%	5.12
	20	100	220	7.18	2.59	222.131	0.97%	5.83	227.997	3.64%	17.47
	20	1000	220	600.60	85.18	221.101	0.50%	496.13	227.997	3.64%	20.64
Average	10			25.54	10.65		3.32%	17.12		7.99%	2.14
	20			302.26	131.01		0.62%	243.73		4.24%	10.41
		100		3.02	1.67		2.34%	2.13		5.81%	2.36
		1000		324.79	139.99		1.60%	258.72		6.42%	10.19
Overall Average				163.90	70.83		1.97%	130.42		6.12%	6.28

time needed for MISOCP-Approx (6.28s vs. 130.42s). As a summary, the four formulations are ranked as follows in terms of computational efficiency: LP-Approx $\succ$ MILP $\succ$ MISOCP-Approx $\succ$ MISOCP.

The number of items M and data points N have both a strong impact on the solution time for each formulation. The solution time of the four formulations is monotone increasing with the M and N. The number of data points N affects more the solution time of the MISOCP, MILP and MISOCP-Approx formulations than it does for the LP-Approx formulation. When N is equal to 1000, the average solution time for MISOCP (resp., MILP and MISOCP-Approx) increases by 106 (resp., 82 and 120) times compared to the time needed for instances with $N = 100$. However, when N increases from 100 to 1000, the solution time for LP-Approx only increases by factor of 3.31 (from 2.36s to 10.19s). The rationale behind this is that the number of binary variables involved in the MISOCP, MILP and MISOCP-Approx formulations is equal to the number of data points N, while the LP-Approx does not include any binary variable. Clearly, the impact of N on the solution time is relatively minor on the LP-Approx formulation, which is the one that scales best for large number of data points.

Table 2 Numerical results with exact reformulation and outer approximation for binary and continuous DRMKP problems: Joint ($T=5$) chance constraint case

Parameters			Binary DRMKP			Continuous DRMKP					
				MISOCP	MILP	MISOCP-Approx			LP-Approx		
$T=5$	M	N	g^*	Time	Time	g^*_{SO}	GAP_{SO}	Time	g^*_{LP}	GAP_{LP}	Time
$\epsilon=0.05, \theta=0.01$	10	100	100	13.54	3.43	104.449	4.45%	8.56	107.993	7.99%	1.89
	10	1000	100	2138.43	427.19	103.205	3.21%	1834.21	107.993	7.99%	125.75
	20	100	215	180.12	43.18	222.107	3.31%	150.87	227.996	6.04%	5.73
	20	1000	–	3600	3600	–	–	3600	–	–	3600
$\epsilon=0.05, \theta=0.03$	10	100	97	35.58	3.23	99.452	2.53%	30.89	107.993	11.33%	1.31
	10	1000	97	2712.12	321.32	99.777	2.86%	2501.05	107.993	11.33%	155.78
	20	100	215	248.13	57.25	220.146	2.39%	204.65	227.996	6.04%	7.30
	20	1000	–	3600	3600	–	–	3600	–	–	3600
$\epsilon=0.10, \theta=0.01$	10	100	100	12.50	3.42	103.504	3.50%	10.12	107.993	7.99%	1.22
	10	1000	100	2481.22	488.44	102.645	2.65%	2012.16	107.993	7.99%	315.69
	20	100	215	200.15	43.70	220.555	2.58%	160.12	227.997	6.05%	1.69
	20	1000	–	3600	3600	–	–	3600	–	–	3600
$\epsilon=0.10, \theta=0.03$	10	100	100	16.91	3.12	103.458	3.46%	10.13	107.993	7.99%	1.42
	10	1000	100	2672.05	459.94	103.349	3.35%	2421.14	107.993	7.99%	309.20
	20	100	215	280.23	41.14	220.101	2.37%	240.12	227.997	6.05%	1.83
	20	1000	–	3600	3600	–	–	3600	–	–	3600
Average	10			1260.29	213.76		3.25%	1103.53		8.83%	114.03
	20			1913.58	1823.16		2.66%	1894.47		6.04%	1802.07
		100		123.40	24.81		3.07%	101.93		7.44%	2.80
		1000		3050.48	2012.11		3.02%	2896.07		8.83%	1913.30
Overall Average				1586.94	1018.46		3.05%	1499.00		7.90%	1014.30

In terms of (approximation) tightness, the optimality gap GAP_{SO} ranges from 0.40% to 4.81% with an overall average of 1.97% across all instances, while the optimality gap GAP_{LP} ranges from 3.64% to 8.00% with an overall average of 6.12%. This demonstrates that the two outer approximation formulations can find high-quality solutions close to the optimal one. Comparing the two approximated approaches, the MISOCP-Approx yields a tighter approximation than that provided by LP-Approx method across all problem instances. The tightness of the MISOCP-Approx is due to the extra second-order cone constraints (43e)-(43h) introduced in $\mathcal{Z}^{CR}_{CSO}$. In summary, while the MISOCP-Approx formulation is tighter than the LP-Approx one (i.e., 1.97% vs 6.12% average optimality gap), LP-Approx is quicker to solve than MISOCP-Approx (6.28 sec. vs. 130.42s average solution time) and scales better.

6.2.2. Results for Joint Chance Constraints Table 2 displays the results for **DRMKP** with joint chance constraint ($T=5$). All binary **DRMKP** models, except for one instance, can be solved to optimality within 46 minutes (2760 seconds). Among the instances solved to optimality, the MISOCP reformulation of the instance with 10 items and 1000 data points takes the longest time to solve, i.e., 2712.12s. For the problem instance with 20 items and 1000 data points, none of the four reformulation and approximation methods proves optimality within the one hour time limit. For the exact reformulation methods, the MILP reformulation solves the problem instance faster than the MISOCP reformulation. On average, the MILP reformulation takes 64% of the time needed to solve the MISOCP reformulation (1018.46s vs. 1586.94s). Excluding the instance not solved to optimality ($M = 20$ and $N = 1000$), the MILP only takes 17% of the time taken by the MISOCP reformulation (157.95s vs. 915.92s). A similar pattern is observed for the solution of the MISOCP and LP outer approximations for the continuous **DRMKP** problems. Omitting the instance not solved to optimality, the LP-Approx solves each problem instance significantly faster than MISOCP-Approx. On average, the LP-Approx method takes 64% of the time needed for the MISOCP-Approx formulation (1014.30s vs. 1499.00). Omitting the instance not solved to optimality, the LP-Approx solution time represents 9.7% of the time needed for MISOCP-Approx (77.40s vs. 789.67s). The overall ranking of the four proposed formulations is: LP-Approx $\succ$ MILP $\succ$ MISOCP-Approx $\succ$ MISOCP, which is consistent with our insights in the individual chance constraint case.

Both the number of items M and data points N significantly impact the solution time of each formulation. The pattern observed for single chance constraints applies to joint chance constraints in the sense that the solution time of the four formulations is monotone increasing with M and N. We note that none of the four formulations is able to solve the instance $M = 20$ and $N = 1000$ to optimality within 3600s, which has a major impact when computing the average solution time and could possibly bias the results. For problem instances with 10 items, when N is equal to 1000, the average solution time for MISOCP (resp., MILP, MISOCP-Approx and LP-Approx) increases by 127 (resp., 128, 146 and 155) times compared to the time needed for instances with $N = 100$. However, the LP-Approx method is still the formulation that scales best for large number of data points, since, for all instances solved, the LP-Approx formulation takes the smallest solution time.

Regarding tightness, the optimality gap GAP_{SO} ranges from 2.37% to 4.45% with an overall average of 3.05% across all instances, while the optimality gap GAP_{LP} ranges from 6.04% to 11.33% with an overall average of 7.90%. This demonstrates that the two outer approximation formulations can find good-quality solutions, close to the optimal one, though the gap ranges is larger compared to the ones obtained in the individual chance constraint case. Comparing the two approximated approaches, similar patterns are extended from the **DRMKP** with individual chance constraint to the joint one. The MISOCP-Approx gives a tighter approximation than that provided by LP-Approx across all instances. While the MISOCP-Approx formulation is tighter than the LP-Approx one (i.e., 3.05% vs 7.90% average optimality gap), LP-Approx is quicker to solve than MISOCP-Approx (1014.30s vs. 1499.00s average solution time) and scales better.

6.3. Selection of Wasserstein Radius via Cross-Validation

The size of Wassserstein radiues θ can have a significant impact on the out-of-sample performance of the DRO solutions. In this subsection, we use the binary DRMKP problem with single chance constraint to numerically illustrate how to choose a proper Wasserstein radius θ via cross-validation. We employ the holdout method (i.e., two-fold cross-validation) to evaluate the out-of-sample performance. We follow a procedure similar to the one used by Esfahani and Kuhn (2018) and proceed as follows:

1. The sample ξ^0 of random weights is generated from a uniform distribution U with a sample size N. We then partition the generated N samples into a training dataset of size N_T and a validation dataset of size $N_V = N - N_T$.

2. Using the training dataset, we solve the binary **DRMKP** problem to optimality with the exact MILP reformulation for a finite number L of candidate radius $\theta \in \{\theta_1, \theta_2, \ldots, \theta_L\}$. For each candidate radius θ_l, we obtain a solution $x_{(\theta_l)}$.

3. Using the validation dataset, we evaluate the out-of-sample performance (i.e., violation probability of the uncertain constraints) of $x_{(\theta_l)}$ via sample average approximation. That is, we apply our solution $x_{(\theta_l)}$ to the validation dataset with sampled weights $\xi^0 \in N_V$, and check if the constraint $\xi_i^{0T} x < b$ is violated or not under sample ξ_i^0, and compute the overall violation probability across samples in the validation set.

4. We repeat the procedure (Steps 1 to 3) K times with independent runs and output the α-percentile of the estimated violation probabilities. The level of α is a user-defined value, which implies that approximately with probability at least α (e.g., 0.90), the solution to the **DRMKP** model from the training dataset is also feasible in the validation dataset.

5. We set the best radius θ^* to the smallest candidate value θ_l such that its α-percentile violation is below the target violation level ϵ (e.g., 0.05). Report $x_{(\theta_l)}$ as the optimal data-driven solution to the **DRMKP** problem.

In our computational experiments, we set $N = 2000$ with $N_T = 1000$ and $N_V = 1000$, the random samples ξ^0 are generated from the uniform distribution $U(5, 15)$. We consider a total of 10 candidate radius $\theta \in \{0.01, 0.02, \ldots, 0.10\}$, and a target risk tolerance level $\epsilon = 0.05$, with a number of $K = 10$ independent runs for each candidate radius. Table 3 displays the numerical results on the selection of a proper Wasserstein radius via cross-validation. We use the notation ρ_k to denote the violation probability in the kth run. We report the 90-percentile of the violation probability for each radius candidate θ_l and denote it by "90-Per".

The results in Table 3 indicate that the best size for the radius of the Wasserstein ambiguity ball is $\theta^* = 0.03$, which is the smallest radius candidate whose 90-percentile violation probability (0.047) is below the target risk level at $\epsilon = 0.05$. Out of the 100 validation tests (10 candidate radius and 10 runs for each), there are only 9 cases when the violation probability exceeds the target risk level of 0.05. The 90-percentiles of violation probability of Wasserstein radius candidates $\theta \in \{0.03, 0.04, 0.06, 0.07, 0.08, 0.09\}$ are below the target risk level. For the binary **DRMKP** models, by choosing the Wasserstein radius properly, the violation probability on the validation dataset is often smaller than the target risk level. This demonstrates the robustness of the proposed distributionally robust multi-item knapsack problem with chance constraints.

Table 3 Choice of Wasserstein Radius Size via Cross-Validation

θ	90-Per	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9	ρ_{10}
0.01	0.051	0.027	0.035	0.047	0.054	0.037	0.051	0.045	0.023	0.035	0.051
0.02	0.050	0.043	0.050	0.038	0.037	0.050	0.014	0.045	0.033	0.055	0.039
0.03	0.047	0.047	0.025	0.039	0.028	0.048	0.028	0.045	0.037	0.016	0.018
0.04	0.044	0.044	0.031	0.038	0.045	0.043	0.019	0.040	0.025	0.019	0.021
0.05	0.051	0.053	0.048	0.038	0.027	0.028	0.051	0.037	0.038	0.031	0.048
0.06	0.039	0.019	0.024	0.025	0.049	0.027	0.015	0.028	0.032	0.039	0.021
0.07	0.026	0.015	0.025	0.012	0.017	0.026	0.016	0.031	0.025	0.011	0.015
0.08	0.040	0.040	0.024	0.034	0.036	0.011	0.044	0.020	0.030	0.023	0.016
0.09	0.043	0.039	0.015	0.031	0.043	0.015	0.013	0.050	0.035	0.017	0.039
0.10	0.052	0.036	0.032	0.055	0.013	0.032	0.052	0.017	0.041	0.045	0.016

7. Conclusion

We investigate distributionally robust chance-constrained programming with data-driven Wasserstein ambiguity sets. We consider two types of Wasserstein uncertainty sets (i.e., with uncertain probabilities and continuum of realizations), two types of distributionally robust chance constraints (i.e., individual and joint), and two types of stochastic functions (i.e., with linear random right-hand side and linear random technology matrix). For the uncertain probabilities cases, we provide new exact mixed-integer linear reformulations for both functional forms. For the continuum of realizations cases, we provide exact mixed-integer second-order cone program (MISOCP) reformulation and a linear program (LP) outer approximation for the linear random right-hand side. We also derive an MISOCP and LP outer approximation for the random technology case. All proposed outer approximations are exact reformulations if the decisions are bounded integer variables. The computational tests conducted on a distributionally robust multi-item knapsack problem (**DRMKP**) with random technology and continuum of realizations attest the scalability of the proposed reformulation/approximation methods. The proposed MILP reformulation with binary variables displays great computational efficiency, while the MISOCP approximation method provides a tight relaxation of the exact reformulation than the LP relaxation. We also discuss how to choose a proper Wasserstesin radius via cross-validation.

References

Blanchet J, Chen L, Zhou XY (2018) Distributionally robust mean-variance portfolio selection with Wasserstein distance Working Paper: Available at Optimization Online.

Boyd S, Vandenberghe L (2004) *Convex optimization* (Cambridge University Press).

Calafiore GC (2007) Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization* 18(3):853–877.

Chen Z, Kuhn D, Wiesemann W (2018) Data-driven chance constrained programs over Wasserstein balls Working Paper: Available at Optimization Online.

Cheng J, Delage E, Lisser A (2014) Distributionally robust stochastic knapsack problem. *SIAM Journal on Optimization* 24(3):1485–1506.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 58(3):595–612.

Duan C, Fang W, Jiang L, Yao L, Liu J (2017) Distributionally robust chance-constrained voltage-concerned dc-opf with Wasserstein metric Working Paper: Available at arXiv: 1706.05538v1.

Erdoğan E, Iyengar G (2006) Ambiguous chance constrained problems and robust optimization. *Mathematical Programming* 107(1-2):37–61.

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

Gao R, Kleywegt AJ (2016) Distributionally robust stochastic optimization with Wasserstein distance, Working Paper: Available at Optimization Online.

Hanasusanto GA, Roitch V, Kuhn D, Wiesemann W (2017) Ambiguous joint chance constraints under mean and dispersion information. *Operations Research* 65(3):751–767.

Hu Z, Hong LJ (2013) Kullback-Leibler divergence constrained distributionally robust optimization Working Paper: Available at Optimization Online.

IBM (2017) IBM big data hub - the four v's of big data URL `http://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg`.

Ji R, Lejeune M (2017) Data-driven optimization of reward-risk ratio measures. *Working Paper: Available at SSRN:* URL `http://dx.doi.org/10.2139/ssrn.2707122`.

Jiang R, Guan Y (2015) Data-driven chance constrained stochastic program. *Mathematical Programming* 158:291–327.

Jiang R, Guan Y, Watson JP (2016) Risk-averse stochastic unit commitment with incomplete information. *IIE Transactions* 48(9):838–854.

McCormick GP (1976) Computability of global solutions to factorable nonconvex programs: Part I – Convex underestimating problems. *Mathematical Programming* 10(1):147–175.

Pflug GC, Pichler A, Wozabal D (2012) The 1/N investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance* 36(2):410–417.

Pflug GC, Pohl M (2017) A review on ambiguity in stochastic portfolio optimization. *Set-Valued and Variational Analysis* 1–25.

Pflug GC, Wozabal D (2007) Ambiguity in portfolio selection. *Quantitative Finance* 7(4):435–442.

Postek K, Den Hertog D, Melenberg B (2016) Computationally tractable counterparts of distributionally robust constraints on risk measures. *SIAM Review* 58(4):603–650.

Wiesemann W, Kuhn D, Sim M (2014) Distributionally robust convex optimization. *Operations Research* 62(6):1358–1376.

Xie W (2018) On distributionally robust chance constrained program with Wasserstein distance. *arXiv preprint arXiv:1806.07418* .

Xie W, Ahmed S (2018) Bicriteria approximation of chance constrained covering problems Working Paper: Available at Optimization Online.

Yang W, Xu H (2016) Distributionally robust chance constraints for non-linear uncertainties. *Mathematical Programming* 155(1-2):231–265.

Zhao C, Guan Y (2018) Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters* 46(2):262–267.

Zymler S, Kuhn D, Rustem B (2013) Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming* 137(1-2):167–198.