



Distributionally robust chance-constrained programs with right-hand side uncertainty under Wasserstein ambiguity

Nam Ho-Nguyen¹ · Fatma Kılınç-Karzan² · Simge Küçükyavuz³ · Dabeen Lee⁴ 

Received: 27 March 2020 / Accepted: 7 December 2020 / Published online: 4 February 2021
© Springer-Verlag GmbH Germany, part of Springer Nature and Mathematical Optimization Society 2021

Abstract

We consider exact deterministic mixed-integer programming (MIP) reformulations of distributionally robust chance-constrained programs (DR-CCP) with random right-hand sides over Wasserstein ambiguity sets. The existing MIP formulations are known to have weak continuous relaxation bounds, and, consequently, for hard instances with small radius, or with large problem sizes, the branch-and-bound based solution processes suffer from large optimality gaps even after hours of computation time. This significantly hinders the practical application of the DR-CCP paradigm. Motivated by these challenges, we conduct a polyhedral study to strengthen these formulations. We reveal several hidden connections between DR-CCP and its nominal counterpart (the sample average approximation), mixing sets, and robust 0–1 programming. By exploiting these connections in combination, we provide an improved formulation and two classes of valid inequalities for DR-CCP. We test the impact of our results on a stochastic transportation problem numerically. Our experiments demonstrate the effectiveness of our approach; in particular our improved formulation and proposed valid inequalities reduce the overall solution times remarkably. Moreover, this allows us to significantly scale up the problem sizes that can be handled in such DR-CCP formulations by reducing the solution times from hours to seconds.

Keywords Distributionally robust · Chance constraints · Mixing sets · Wasserstein ambiguity

Mathematics Subject Classification 90-08 · 90-10 · 90C10 · 90C11 · 90C15 · 90C17 · 90C27 · 90C57

1 Introduction

We consider the following chance-constrained program (CCP)

Extended author information available on the last page of the article

$$\begin{aligned}
 \min_{\mathbf{x}} \quad & \mathbf{c}^\top \mathbf{x} && \text{(CCP)} \\
 \text{s.t.} \quad & \mathbb{P}^*[\boldsymbol{\xi} \notin \mathcal{S}(\mathbf{x})] \leq \epsilon, \\
 & \mathbf{x} \in \mathcal{X},
 \end{aligned}$$

where $\mathbf{c} \in \mathbb{R}^L$ is a cost vector, $\mathcal{X} \subset \mathbb{R}^L$ is a compact domain for the decision variables \mathbf{x} , $\mathcal{S}(\mathbf{x}) \subseteq \mathbb{R}^K$ is a decision-dependent safety set, $\boldsymbol{\xi} \in \mathbb{R}^K$ is a random variable with distribution \mathbb{P}^* , and $\epsilon \in (0, 1)$ is the risk tolerance for the random variable $\boldsymbol{\xi}$ falling outside the safety set $\mathcal{S}(\mathbf{x})$.

CCP is one of the most common models to handle uncertainty in optimization. Nevertheless, in practice, the distribution \mathbb{P}^* in the chance constraint in (CCP) is often unavailable to the optimizer. Instead, independent and identically distributed (i.i.d.) samples $\{\boldsymbol{\xi}_i\}_{i \in [N]}$, where $[N] := \{1, \dots, N\}$, are drawn from \mathbb{P}^* , and \mathbb{P}^* is approximated using the empirical distribution \mathbb{P}_N on these samples. Such an approach is known as the *sample average approximation* (SAA) of (CCP). Note that evaluating $\mathbb{P}^*[\boldsymbol{\xi} \notin \mathcal{S}(\mathbf{x})]$ exactly is often difficult even when \mathbb{P}^* is available. Consequently, the SAA approach is often employed whenever the computation of $\mathbb{P}^*[\boldsymbol{\xi} \notin \mathcal{S}(\mathbf{x})]$ is expensive, even if \mathbb{P}^* is available. The SAA formulation of (CCP) is

$$\begin{aligned}
 \min_{\mathbf{x}} \quad & \mathbf{c}^\top \mathbf{x} && \text{(SAA)} \\
 \text{s.t.} \quad & \frac{1}{N} \sum_{i \in [N]} \mathbf{1}(\boldsymbol{\xi}_i \notin \mathcal{S}(\mathbf{x})) \leq \epsilon, \\
 & \mathbf{x} \in \mathcal{X},
 \end{aligned}$$

where $\mathbf{1}(\cdot)$ is the indicator function. For certain forms of safety sets $\mathcal{S}(\cdot)$, (SAA) can be reformulated as a mixed-integer program (MIP) and thus off-the-shelf optimization solvers can be used to solve it.

While there are statistical guarantees for using (SAA) to approximate (CCP) [5,7,24], the out-of-sample performance of the solution from (SAA) is quite sensitive to the specific sample $\{\boldsymbol{\xi}_i\}_{i \in [N]}$, and can result in high variance, particularly for small N . In order to remedy this and regularize the out-of-sample performance of (SAA), one can solve a *distributionally robust chance-constrained program*:

$$\begin{aligned}
 \min_{\mathbf{x}} \quad & \mathbf{c}^\top \mathbf{x} && \text{(DR-CCP)} \\
 \text{s.t.} \quad & \sup_{\mathbb{P} \in \mathcal{F}_N(\theta)} \mathbb{P}[\boldsymbol{\xi} \notin \mathcal{S}(\mathbf{x})] \leq \epsilon, \\
 & \mathbf{x} \in \mathcal{X},
 \end{aligned}$$

where $\mathcal{F}_N(\theta)$ is an ambiguity set of distributions on \mathbb{R}^K that contains the empirical distribution \mathbb{P}_N , and θ is a parameter that governs the size of the ambiguity set, and thereby the conservatism of (DR-CCP). For a recent comprehensive survey on distributionally robust optimization problems, their properties, and exact and approximate methods to solve them, we refer the reader to Rahimian and Mehrotra [30] and references therein.

Several types of ambiguity sets of probability distributions, such as those based on moments, ϕ -divergences, unimodality, or support have been studied in the literature; see e.g., [6,9,12,15,19,33]. More recently, the *Wasserstein* ambiguity set, that is, the Wasserstein distance ball of radius θ around the empirical distribution \mathbb{P}_N , is shown to possess particularly attractive statistical properties. The dual representation for the worst-case probability $\mathbb{P}[\xi \notin \mathcal{S}(\mathbf{x})]$ under the Wasserstein ambiguity set $\mathbb{P} \in \mathcal{F}_N(\theta)$ is given in [4,10,27]. Various studies [8,13,32] exploit this dual representation to give a deterministic non-convex reformulation of (DR-CCP). See also Hota et al. [13] for reformulations based on the conditional value-at-risk (CVaR) inner approximation of (DR-CCP) for several different types of safety sets.

For common linear forms of safety sets $\mathcal{S}(\cdot)$, Chen et al. [8] and Xie [32] show that (DR-CCP) under Wasserstein ambiguity sets admits a MIP reformulation. Ji and Lejeune [14] also explore MIP formulations of (DR-CCP) under Wasserstein ambiguity. However, they impose additional structure on the support of ξ , thus their formulations are different from Chen et al. [8] and Xie [32]. Such MIP reformulations pave the path of using standard optimization solvers to formulate and solve these problems, however, according to [8,32] these MIP reformulations are difficult to solve in certain cases. Even for relatively small size problems with only a few hundred samples and small radii, the resulting formulations still have a large optimality gap even after an hour of computation time with a commercial solver.

1.1 Contributions

Motivated by these computational challenges, our focus in this paper is on developing effective methods to solve the exact reformulation of (DR-CCP) for such hard cases. In particular, we closely examine the MIP reformulations of (DR-CCP) with random right-hand side uncertainty under Wasserstein ambiguity sets from Chen et al. [8] and Xie [32]. We present theoretical results that have strong computational impact in solving (DR-CCP).

We first exploit the close relationship between (DR-CCP) over Wasserstein ambiguity sets and its nominal counterpart (SAA) to identify an implied mixing set with a cardinality constraint over the existing binary variables in the (DR-CCP) reformulation, which enables the use of existing inequalities for such a set. We use the established technique of *quantile strengthening* [23,25] to significantly reduce the big- M constants of this mixing set, and then show how to adapt these to (DR-CCP). Our result also shows that existing mixing inequalities for (SAA) can be readily applied to (DR-CCP).

We further analyze the formulation with reduced coefficients obtained from the quantile strengthening. We exploit the conditional value-at-risk (CVaR) interpretation for (DR-CCP) described by Xie [32] to bound key variables in the formulation. Using these, we show that we can further improve our new formulation for (DR-CCP) by eliminating a significant portion of constraints. Furthermore, the improved bounds we establish on variables allow us to identify a specific substructure in our improved formulation which arises in robust 0–1 programming [2]. Consequently, we show that we can directly use results from Atamtürk [2] to describe a class of valid inequalities for our improved formulation.

We then assess the computational impact of the improved formulation and the new valid inequalities on solving (DR-CCP) on a class of stochastic transportation problems. We observe that the improved formulation uniformly reduces solution times by at least an order of magnitude for any radius, number of original decision variables and number of scenarios. In the difficult small radius regime for $N = 100$ scenarios, the original formulation cannot verify optimality within 1 h, whereas our formulation solves within seconds. For the most difficult instances with small Wasserstein radius and $N = 3000$ scenarios, the improved formulation results in less than 0.8% average optimality gap, whereas the original formulation cannot even find a feasible solution for any of the tested instances.

To the best of our knowledge, our work is the first that examines these connections between traditional SAA, mixing sets, and robust 0–1 programming in the context of distributionally robust chance constraints, and demonstrates the huge computational effectiveness of these approaches for the difficult instances. It is interesting to note that distributionally robust optimization is a paradigm of modeling uncertainty that does not require a complete knowledge of the distribution as in (SAA), but is also less conservative than robust optimization that considers the worst-case realizations of uncertain parameters, without any knowledge of their joint distribution. Nevertheless, our analysis and computational results show that the formulations for distributionally robust optimization can be significantly improved by employing its connections to both SAA and robust optimization.

1.2 Outline

In Sect. 2, we provide a formal problem description and elaborate on the connections between the existing MIP models for (DR-CCP) and the safety set. In Sect. 3, we explore the connection between (SAA) and (DR-CCP). We describe the mixing structure in (SAA) and demonstrate its application to (DR-CCP). In Sect. 4, we employ the CVaR representation of (DR-CCP) to reveal the substructure from robust 0–1 programming that is hidden in MIP reformulations of (DR-CCP). Using results from Atamtürk [2], we provide valid inequalities resulting from this substructure. In Sect. 5, we report our computational experience with the improved formulation and the proposed inequalities on a class of stochastic transportation problems.

2 Problem formulation

We consider Wasserstein ambiguity sets $\mathcal{F}_N(\theta)$ defined as the θ -radius Wasserstein ball of distributions on \mathbb{R}^K around the empirical distribution \mathbb{P}_N . We will use the l -Wasserstein distance, based on a norm $\|\cdot\|$, between two distributions \mathbb{P} and \mathbb{P}' . This is defined as follows:

$$d_W(\mathbb{P}, \mathbb{P}') := \inf_{\Pi} \left\{ \mathbb{E}_{(\xi, \xi') \sim \Pi} [\|\xi - \xi'\|] : \Pi \text{ has marginal distributions } \mathbb{P}, \mathbb{P}' \right\}.$$

Then, the Wasserstein ambiguity set is

$$\mathcal{F}_N(\theta) := \{\mathbb{P} : d_W(\mathbb{P}_N, \mathbb{P}) \leq \theta\}.$$

Given a decision $\mathbf{x} \in \mathcal{X}$ and random realization $\xi \in \mathbb{R}^K$, the *distance from ξ to the unsafe set* is

$$\text{dist}(\xi, \mathcal{S}(\mathbf{x})) := \inf_{\xi' \in \mathbb{R}^K} \{ \|\xi - \xi'\| : \xi' \notin \mathcal{S}(\mathbf{x}) \}. \tag{1}$$

Throughout Sects. 2, 3 and 4, we assume that the sample $\{\xi_i\}_{i \in [N]}$, the risk tolerance $\epsilon \in (0, 1)$ and the radius $\theta > 0$ are fixed. As short-hand notation, we denote the feasible regions of (SAA) and (DR-CCP) as follows:

$$\mathcal{X}_{\text{SAA}}(\mathcal{S}) := \left\{ \mathbf{x} \in \mathcal{X} : \frac{1}{N} \sum_{i \in [N]} \mathbf{1}(\xi_i \notin \mathcal{S}(\mathbf{x})) \leq \epsilon \right\}, \tag{2a}$$

$$\mathcal{X}_{\text{DR}}(\mathcal{S}) := \left\{ \mathbf{x} \in \mathcal{X} : \sup_{\mathbb{P} \in \mathcal{F}_N(\theta)} \mathbb{P}[\xi \notin \mathcal{S}(\mathbf{x})] \leq \epsilon \right\}. \tag{2b}$$

Note that here the dependence on the safety set function \mathcal{S} is made explicit, since the relationship between existing formulations and our new valid inequalities depends on the safety set.

Using tools from duality theory for Wasserstein distributional robustness [4,10], Chen et al. [8] and Xie [32] give an extended formulation for the distributionally robust chance constraint in (DR-CCP). Specifically, it was shown in [8, Theorem 3] (see also [32, Proposition 1]) that when $\mathcal{S}(\mathbf{x})$ is open for each $\mathbf{x} \in \mathcal{X}$ and $\theta > 0$, we have

$$\mathcal{X}_{\text{DR}}(\mathcal{S}) = \left\{ \mathbf{x} \in \mathcal{X} : \begin{array}{l} \exists t \geq 0, \mathbf{r} \geq \mathbf{0}, \\ \text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) \geq t - r_i, i \in [N], \\ \epsilon t \geq \theta + \frac{1}{N} \sum_{i \in [N]} r_i \end{array} \right\}. \tag{3}$$

(Note that a similar formulation holds when $\theta = 0$, but we need to make the restriction $t > 0$.) Therefore, the ability to model (DR-CCP) depends on the ability to model the constraints $\text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) \geq t - r_i$. For a given set $\mathcal{S}(\mathbf{x})$, let $\text{int } \mathcal{S}(\mathbf{x})$ denote its interior and $\text{cl } \mathcal{S}(\mathbf{x})$ denote its closure. Also, note that Gao and Kleywegt [10, Proposition 3] show

$$\sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}[\xi \notin \text{int } \mathcal{S}(\mathbf{x})] = \sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}[\xi \notin \mathcal{S}(\mathbf{x})] = \sup_{\mathbb{P} \in \mathcal{F}(\theta)} \mathbb{P}[\xi \notin \text{cl } \mathcal{S}(\mathbf{x})],$$

and $\text{dist}(\xi, \text{int } \mathcal{S}(\mathbf{x})) = \text{dist}(\xi, \mathcal{S}(\mathbf{x})) = \text{dist}(\xi, \text{cl } \mathcal{S}(\mathbf{x}))$. Therefore, (3) holds regardless of whether $\mathcal{S}(\mathbf{x})$ is open or closed. With this in mind, in what follows, we express $\mathcal{S}(\mathbf{x})$ as an open set, for convenience.

The classical literature on CCP typically considers two types of safety sets which are defined by linear inequalities and are known to admit exact MIP reformulations: *individual chance constraints with left-hand side (LHS) uncertainty* and *joint chance*

constraints with right-hand side (RHS) uncertainty. In this paper, we consider only joint chance constraints with RHS uncertainty, which have safety set and distance function given by

$$\mathcal{S}(\mathbf{x}) := \left\{ \boldsymbol{\xi} : \mathbf{b}_p^\top \boldsymbol{\xi} + d_p - \mathbf{a}_p^\top \mathbf{x} > 0, p \in [P] \right\}, \tag{4a}$$

$$\text{dist}(\boldsymbol{\xi}, \mathcal{S}(\mathbf{x})) = \max \left\{ 0, \min_{p \in [P]} \frac{\mathbf{b}_p^\top \boldsymbol{\xi} + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right\}, \tag{4b}$$

for given $\mathbf{a}_p \in \mathbb{R}^K$, $\mathbf{b}_p \in \mathbb{R}^L$ and $d_p \in \mathbb{R}$ for all $p \in [P]$ where $\|\cdot\|_*$ is the dual norm. Chen et al. [8, Proposition 2] show that, in this case, (DR-CCP) can be reformulated as

$$\min_{\mathbf{z}, \mathbf{r}, t, \mathbf{x}} \mathbf{c}^\top \mathbf{x} \tag{5a}$$

$$\text{s.t. } \mathbf{z} \in \{0, 1\}^N, t \geq 0, \mathbf{r} \geq \mathbf{0}, \mathbf{x} \in \mathcal{X}, \tag{5b}$$

$$\epsilon t \geq \theta + \frac{1}{N} \sum_{i \in [N]} r_i, \tag{5c}$$

$$M(1 - z_i) \geq t - r_i, \quad i \in [N], \tag{5d}$$

$$\frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} + Mz_i \geq t - r_i, \quad i \in [N], p \in [P], \tag{5e}$$

where M is a sufficiently large positive constant. In particular, for joint chance constraints with RHS uncertainty, we have

$$\mathcal{X}_{\text{DR}}(\mathcal{S}) = \{\mathbf{x} \in \mathcal{X} : (5b)-(5e)\}. \tag{6}$$

3 Connection with the nominal chance constraint and mixing sets

Our first idea to strengthen the MIP reformulations of distributionally robust CCPs originates from the following simple observation between the relation of the empirical probability distribution and the Wasserstein ambiguity set.

When the radius θ of the Wasserstein ambiguity set $\mathcal{F}_N(\theta)$ is 0, (DR-CCP) coincides with (SAA) since $\mathcal{F}_N(0) = \{\mathbb{P}_N\}$. In general, as $\mathcal{F}_N(0) \subseteq \mathcal{F}_N(\theta)$ for any $\theta \geq 0$, (SAA) is a relaxation of (DR-CCP), i.e., we have

$$\mathcal{X}_{\text{DR}}(\mathcal{S}) \subseteq \mathcal{X}_{\text{SAA}}(\mathcal{S}).$$

When the safety set is defined as $\mathcal{S}(\mathbf{x}) = \{\boldsymbol{\xi} : s(\mathbf{x}, \boldsymbol{\xi}) \geq 0\}$ for a continuous function $s(\cdot)$, Ruszczyński [31] shows that (SAA) can be reformulated as the following MIP:

$$\min_{\mathbf{z}, \mathbf{r}, t, \mathbf{x}} \mathbf{c}^\top \mathbf{x} \tag{7a}$$

$$\text{s.t. } \mathbf{z} \in \{0, 1\}^N, \mathbf{x} \in \mathcal{X}, \tag{7b}$$

$$\sum_{i \in [N]} z_i \leq \lfloor \epsilon N \rfloor, \tag{7c}$$

$$s(\mathbf{x}, \xi_i) + Mz_i \geq 0, \quad i \in [N], \tag{7d}$$

where M is a sufficiently large positive constant. Inequalities (7c) and (7d) are often referred to as the *knapsack (or cardinality) constraint* and the *big- M constraints*, respectively. Then, when the safety set $\mathcal{S}(\mathbf{x})$ is given by $\mathcal{S}(\mathbf{x}) = \{\xi : s(\mathbf{x}, \xi) > 0\}$, (7) provides a relaxation of (DR-CCP). Therefore, by solving (7), one can provide a lower bound on the optimum value of (DR-CCP). Note that the safety set (4) can be written in this form by defining $s(\cdot)$ appropriately.

The formulation (7) for (SAA) is well-studied in the literature, and many classes of valid inequalities for the formulation have been developed; see e.g., [1,16,17,20,22,23,25,34,35]. In fact, we develop a more direct connection between formulation (7) and (DR-CCP) so that we can apply techniques for solving (7) directly to (DR-CCP). It is clear that inequalities of the form (7c)–(7d) can be added to the MIP formulation (5) of (DR-CCP) to obtain a stronger formulation. In turn, this implies that the reformulation can be further strengthened by the inequalities developed for strengthening (7). If we do this naively, we would add (7c)–(7d) to (5) with *new* binary variables $\mathbf{z}' \in \{0, 1\}^N$. Our first key result is that *the same binary variables* \mathbf{z} from the MIP formulation (5) can be used to define (7c)–(7d). This then means that we can strengthen these formulations *without* adding any additional binary variables.

3.1 Strengthening the formulation by the nominal chance constraint

We now verify that the SAA inequalities for the joint chance constraint of (2b) can be used to strengthen the formulation (5). Given $\mathcal{S}(\mathbf{x})$ defined by (4), consider the following MIP formulation

$$\min_{\mathbf{x}, \mathbf{z}, \mathbf{r}, t} \mathbf{c}^\top \mathbf{x} \tag{8a}$$

$$\text{s.t. } (\mathbf{z}, \mathbf{r}, t, \mathbf{x}) \text{ satisfies } (5b)–(5e), \tag{8b}$$

$$\sum_{i \in [N]} z_i \leq \lfloor \epsilon N \rfloor, \tag{8c}$$

$$\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} + Mz_i \geq 0, \quad i \in [N], p \in [P], \tag{8d}$$

where M is a sufficiently large positive constant. (Note that in this formulation, the inequality (8c) is equivalent to (7c).) We can write constraints (8d) in the form (7d): individually $s_p(\mathbf{x}, \xi_i) := \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*}$ or jointly $s(\mathbf{x}, \xi) := \min_{p \in [P]} \left\{ \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right\}$. More importantly, (8c) and (8d) share the same set of binary

variables as the other constraints in (5). We next argue that this MIP formulation (8) is an exact reformulation of (DR-CCP) with a safety set $\mathcal{S}(\mathbf{x})$ from (4).

Theorem 1 *When the safety set \mathcal{S} is defined as in (4), formulation (8) is an exact reformulation of (DR-CCP), i.e.,*

$$\mathcal{X}_{\text{DR}}(\mathcal{S}) = \{\mathbf{x} \in \mathcal{X} : (8b)-(8d)\}. \tag{9}$$

Proof By (6), it is sufficient to show that

$$\{\mathbf{x} \in \mathcal{X} : (5b)-(5e)\} = \{\mathbf{x} \in \mathcal{X} : (8b)-(8d)\}.$$

By (8b), we know that the set on the right-hand side is contained in the set on the left-hand side. Let us show that the reverse direction also holds. To this end, take a vector $\mathbf{x} \in \mathcal{X}$ satisfying (5b)–(5e) with some $\mathbf{z}, \mathbf{r}, t$. For ease of notation, we define $s(\mathbf{x}, \xi_i) := \min_{p \in [P]} \left\{ \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right\}$. We claim that $(\mathbf{x}, \bar{\mathbf{z}}, \mathbf{r}, t)$ satisfies (8b)–(8d) where $\bar{\mathbf{z}} \in \{0, 1\}^N$ is a vector such that $\bar{z}_i = 1$ if and only if $s(\mathbf{x}, \xi_i) < 0$ for all $i \in [N]$. Since M is sufficiently large so that $s(\mathbf{x}, \xi_i) + M \geq 0$, it is clear that $\mathbf{x}, \bar{\mathbf{z}}$ satisfy (8d). Next, we observe that $(\mathbf{x}, \bar{\mathbf{z}}, \mathbf{r}, t)$ satisfies (5d) and (5e), which can be equivalently rewritten as

$$\min \{s(\mathbf{x}, \xi_i) + Mz_i, M(1 - z_i)\} \geq t - r_i, \quad i \in [N].$$

By the choice of $\bar{\mathbf{z}}$ and M , we have $\min \{s(\mathbf{x}, \xi_i) + M\bar{z}_i, M(1 - \bar{z}_i)\}$ is equal to 0 if $s(\mathbf{x}, \xi_i) < 0$, and it is equal to $s(\mathbf{x}, \xi_i)$ otherwise. Hence,

$$\min \{s(\mathbf{x}, \xi_i) + M\bar{z}_i, M(1 - \bar{z}_i)\} \geq \min \{s(\mathbf{x}, \xi_i) + Mz_i, M(1 - z_i)\}$$

for any $z_i \in \{0, 1\}$. This implies that $(\mathbf{x}, \bar{\mathbf{z}}, \mathbf{r}, t)$ satisfies (5d) and (5e) as they are satisfied already by $(\mathbf{x}, \mathbf{z}, \mathbf{r}, t)$. To finish the proof, it remains to show that $\bar{\mathbf{z}}$ satisfies (8c). Since $\theta > 0$, we obtain from (5c) that $t > 0$. This implies that $\frac{t_i}{t} \geq 0$. We claim that $\frac{t_i}{t} \geq \bar{z}_i$ for all $i \in [N]$. As $\frac{t_i}{t} \geq 0$, we have $\frac{t_i}{t} \geq \bar{z}_i$ holds when $\bar{z}_i = 0$. When $\bar{z}_i = 1$, by rearranging (5d) we get $\frac{t_i}{t} \geq 1 = \bar{z}_i$. Since $\frac{t_i}{t} \geq \bar{z}_i$ for all $i \in [N]$, it follows from (5c) that $\epsilon N \geq \sum_{i \in [N]} \frac{t_i}{t} \geq \sum_{i \in [N]} \bar{z}_i$, implying in turn that $\sum_{i \in [N]} \bar{z}_i \leq \lfloor \epsilon N \rfloor$, as required. \square

Remark 1 Chen et al. [8] argue that there is a finite value of M for the validity of (5). Essentially, we need to choose an appropriate value of M so that (5d)–(5e) correctly represent the constraint $\text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) \geq t - r_i$. When $\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x} < 0$ for some $p \in [P]$, we have $\text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) = 0$ by (4) and thus $\text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) \geq t - r_i$ becomes $0 \geq t - r_i$. As long as $\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} + M \geq 0$ holds for all $\mathbf{x} \in \mathcal{X}$, (5d)–(5e) can capture this situation by setting $z_i = 1$. Similarly, (5d)–(5e) can represent the case when $\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x} \geq 0$ for all $p \in [P]$ if $M \geq \mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}$ for all $\mathbf{x} \in \mathcal{X}, p \in [P]$. So, setting

$$M := \max_{\mathbf{x} \in \mathcal{X}, p \in [P]} \left\{ \frac{|\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}|}{\|\mathbf{b}_p\|_*} \right\} \tag{10}$$

ensures that the constraints (8d) are indeed valid. □

3.2 Mixing substructure

Although (8) is already stronger than (5) due to the additional constraints (8d)–(8c), we can further strengthen this formulation by adding more valid inequalities originating from these constraints. To this end, for any fixed p , the constraints (8d)–(8c) give rise to the following substructure:

$$Q_p := \left\{ (\mathbf{x}, \mathbf{z}) \in \mathcal{X} \times \{0, 1\}^N : \begin{array}{l} s_p(\mathbf{x}, \boldsymbol{\xi}_i) + Mz_i \geq 0, \quad i \in [N], \\ \sum_{i \in [N]} z_i \leq \lfloor \epsilon N \rfloor \end{array} \right\}. \tag{11}$$

So, one can generate valid inequalities for the formulation (8) by finding inequalities of the form $\boldsymbol{\mu}^\top \mathbf{x} + \boldsymbol{\pi}^\top \mathbf{z} \geq \beta$ that are valid for the mixed-integer set Q_p in (11). Luedtke et al. [25] and Luedtke [23] introduce a procedure of generating inequalities that are valid for Q_p . In order to make our paper self-contained, we next explain this procedure.

Given a fixed linear function $\boldsymbol{\mu}^\top \mathbf{x}$, we solve the following single scenario subproblem for each scenario $i \in [N]$

$$\bar{h}_i(\boldsymbol{\mu}) := \min \left\{ \boldsymbol{\mu}^\top \mathbf{x} : s_p(\mathbf{x}, \boldsymbol{\xi}_i) \geq 0, \mathbf{x} \in \bar{\mathcal{X}} \right\}, \tag{12}$$

where $\mathcal{X} \subseteq \bar{\mathcal{X}} \subseteq \mathbb{R}^L$. Then $\boldsymbol{\mu}^\top \mathbf{x} \geq \bar{h}_i(\boldsymbol{\mu})$ holds for $(\mathbf{x}, \mathbf{z}) \in Q_p$ with $z_i = 0$. Having computed the values $\bar{h}_i(\boldsymbol{\mu})$ for $i \in [N]$, we sort them in non-decreasing order. Without loss of generality, we may assume that

$$\bar{h}_N(\boldsymbol{\mu}) \geq \bar{h}_{N-1}(\boldsymbol{\mu}) \geq \dots \geq \bar{h}_1(\boldsymbol{\mu}).$$

For ease of notation, let

$$k := \lfloor \epsilon N \rfloor.$$

Notice that because $\sum_{i \in [N]} z_i \leq k$ is also enforced in Q_p , by the pigeonhole principle there must exist $i \in \{N - k, N - k + 1, \dots, N\}$ with $z_i = 0$. In turn, this implies that $\boldsymbol{\mu}^\top \mathbf{x} \geq \bar{h}_{N-k}(\boldsymbol{\mu})$ because $\bar{h}_i(\boldsymbol{\mu}) \geq \bar{h}_{N-k}(\boldsymbol{\mu})$ for all $i \geq N - k$. In summary, we have just argued that $\boldsymbol{\mu}^\top \mathbf{x} \geq \bar{h}_i(\boldsymbol{\mu})$ holds if $z_i = 0$ and that $\boldsymbol{\mu}^\top \mathbf{x} \geq \bar{h}_{N-k}(\boldsymbol{\mu})$ is satisfied always, in particular, when $z_i = 1$ for $i \in [N]$. Equivalently,

$$\boldsymbol{\mu}^\top \mathbf{x} + (\bar{h}_i(\boldsymbol{\mu}) - \bar{h}_{N-k}(\boldsymbol{\mu})) z_i \geq \bar{h}_i(\boldsymbol{\mu}) \tag{13}$$

is valid. In fact, inequalities (13) for $i \leq N - k$ are redundant because $\boldsymbol{\mu}^\top \mathbf{x} \geq \bar{h}_i(\boldsymbol{\mu})$ is implied by $\boldsymbol{\mu}^\top \mathbf{x} \geq \bar{h}_{N-k}(\boldsymbol{\mu})$ if $i \leq N - k$. Because inequalities (13) share a common linear function $\boldsymbol{\mu}^\top \mathbf{x}$ but each one has a distinct integer variable, the *mixing procedure* of Günlük and Pochet [11], can be applied to obtain stronger inequalities. For any $J = \{j_1, \dots, j_\ell\}$ with $N \geq j_1 \geq \dots \geq j_\ell \geq N - k + 1$, the *mixing inequality* derived from J and (13) is

$$\boldsymbol{\mu}^\top \mathbf{x} + \sum_{i \in [\ell]} (\bar{h}_{j_i}(\boldsymbol{\mu}) - \bar{h}_{j_{i+1}}(\boldsymbol{\mu})) z_{j_i} \geq \bar{h}_{j_1}(\boldsymbol{\mu}), \tag{14}$$

where $j_{\ell+1} := N - k$. The mixing inequalities of the form (14) are equivalent to the *star inequalities* by Atamtürk et al. [3]. The inequalities (14) are the strongest possible ones that can be generated from (13) in that the convex hull of solutions $(\mathbf{x}, \mathbf{z}) \in \mathbb{R}^L \times \{0, 1\}^N$ satisfying (13) is described by (14) [3, 11, 16].

Consider any $p \in [P]$. For $s_p(\mathbf{x}, \boldsymbol{\xi}_i) := \frac{1}{\|\mathbf{b}_p\|_*} (\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x})$ as in (8), we can take $-\frac{1}{\|\mathbf{b}_p\|_*} \mathbf{a}_p$ for $\boldsymbol{\mu}$. For this choice of $\boldsymbol{\mu}$, the value of $\bar{h}_i(\boldsymbol{\mu})$ from the single scenario subproblem (12) with $\bar{\mathcal{X}} = \mathbb{R}^L$ is precisely

$$\min_{\mathbf{x}} \left\{ -\frac{\mathbf{a}_p^\top}{\|\mathbf{b}_p\|_*} \mathbf{x} : \frac{1}{\|\mathbf{b}_p\|_*} (\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}) \geq 0 \right\} = -\frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p}{\|\mathbf{b}_p\|_*}.$$

So, assuming $-\mathbf{b}_p^\top \boldsymbol{\xi}_N \geq \dots \geq -\mathbf{b}_p^\top \boldsymbol{\xi}_1$ and letting

$$q_p := -\mathbf{b}_p^\top \boldsymbol{\xi}_{N-k}$$

(q_p is the $(k + 1)$ -th largest value), the inequalities $\boldsymbol{\mu}^\top \mathbf{x} \geq \bar{h}_{N-k}(\boldsymbol{\mu})$ and (13) for $\boldsymbol{\mu} = -\frac{1}{\|\mathbf{b}_p\|_*} \mathbf{a}_p$ correspond to

$$\frac{-q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \geq 0, \tag{15a}$$

$$\frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} + \frac{-\mathbf{b}_p^\top \boldsymbol{\xi}_i - q_p}{\|\mathbf{b}_p\|_*} z_i \geq 0, \quad i \in [N]. \tag{15b}$$

Moreover, the mixing inequalities (14) obtained from (13) have the following form:

$$\frac{\mathbf{b}_p^\top \boldsymbol{\xi}_{j_1} + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} + \sum_{i \in [\ell]} \frac{-\mathbf{b}_p^\top \boldsymbol{\xi}_{j_i} + \mathbf{b}_p^\top \boldsymbol{\xi}_{j_{i+1}}}{\|\mathbf{b}_p\|_*} z_{j_i} \geq 0 \tag{16}$$

where $N \geq j_1 \geq \dots \geq j_\ell \geq N - k + 1$ and $j_{\ell+1} := N - k$.

The number of mixing inequalities (16) is exponential, but they can be separated in $O(N \log N)$ time [11, 16].

Notice that the inequalities (13)–(16) are *big-M-free*; the coefficients of the binary variables and the right-hand sides depend only on $s_p(\mathbf{x}, \boldsymbol{\xi}_i)$. This is important for practical purposes. In particular, when M is larger than $\frac{-\mathbf{b}_p^\top \boldsymbol{\xi}_i - q_p}{\|\mathbf{b}_p\|_*}$, the constraints (8d) are dominated by the inequalities (15b), and thus, by (16).

3.3 Reducing big-M values

In Sect. 3.2, we argued that letting q_p be the $(k + 1)$ -th largest value amongst $\{-\mathbf{b}_p^\top \boldsymbol{\xi}_i\}_{i \in [N]}$, the inequalities (15) are valid for (8), and based on this we generate the mixing inequalities (16). In fact, we can replace the big- M in (5e) with its strengthened version from (15b) to obtain a new formulation:

$$\min_{\mathbf{z}, \mathbf{r}, t, \mathbf{x}} \mathbf{c}^\top \mathbf{x} \tag{17a}$$

$$\text{s.t. } (\mathbf{z}, \mathbf{r}, t, \mathbf{x}) \text{ satisfies (5b)–(5d) and (5c),} \tag{17b}$$

$$\frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} + \frac{-\mathbf{b}_p^\top \boldsymbol{\xi}_i - q_p}{\|\mathbf{b}_p\|_*} z_i \geq t - r_i, \quad i \in [N], p \in [P]. \tag{17c}$$

Theorem 2 *Formulation (17) is an exact reformulation of (DR-CCP) where the safety set is given by (4).*

Proof By Theorem 1, (5) with (8c) is an exact reformulation. Hence, we need to argue that reducing the big- M value in (5e) to obtain (17c) keeps the formulation valid. To this end, it suffices to argue that constraints (5d) and (17c) correctly represent $\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x})) \geq t - r_i$ for $i \in [N]$. Let $i \in [N]$. If $\min_{p \in [P]} \left\{ \frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right\} \geq 0$, we can set $z_i = 0$ and we obtain $\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x})) = \min_{p \in [P]} \left\{ \frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right\}$. In this case, (5d) becomes redundant and (17c) represents $\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x})) \geq t - r_i$. On the other hand, if we set $z_i = 1$, then (5d) is $0 \geq t - r_i$ and (17c) becomes redundant. However, $\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x})) \geq t - r_i$ is less restrictive on $(\mathbf{x}, \mathbf{r}, t)$ than $0 \geq t - r_i$, so at optimality, there is at least one solution such that $z_i = 0$ whenever $\min_{p \in [P]} \left\{ \frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right\} \geq 0$, hence $\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x})) \geq t - r_i$ is correctly represented.

If $\min_{p \in [P]} \left\{ \frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \right\} < 0$, then $\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x})) = 0$ and we must have $z_i = 1$ by (15b). Since $z_i = 1$, (17c) and (5d) respectively become

$$t - r_i \leq \frac{\mathbf{b}_p^\top \boldsymbol{\xi}_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} + \frac{-\mathbf{b}_p^\top \boldsymbol{\xi}_i - q_p}{\|\mathbf{b}_p\|_*} = \frac{-q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*}$$

$$t - r_i \leq 0.$$

By (15a) we know that $\frac{-q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \geq 0$ so the first inequality is weaker than the second, hence we have $\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x})) = 0 \geq t - r_i$ as required. \square

In contrast to (5e), the inequality (17c) is *big-M-free*. During our computational study summarized in Sect. 5, we observed that the coefficients of the variables \mathbf{z} in (17c) are significantly smaller than the big- M value computed from Remark 1.

4 Improved formulation and valid inequalities from robust 0–1 programming

In this section, we study, in closer detail, the set of constraints from (17c). By appealing to the conditional value-at-risk interpretation for (DR-CCP), we prove a bound on t . Consequently, we uncover another hidden structure related to robust 0–1 programming with a budget uncertainty set. This allows us to derive a new class of valid inequalities.

For a given $p \in [P]$, recall the definition of q_p being the $(k + 1)$ -th largest value amongst $\left\{ -\mathbf{b}_p^\top \boldsymbol{\xi}_i \right\}_{i \in [N]}$ where $k := \lfloor \epsilon N \rfloor$. We also define $h_{i,p} := \frac{-\mathbf{b}_p^\top \boldsymbol{\xi}_i - q_p}{\|\mathbf{b}_p\|_*}$ and $u_p := \frac{-q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} - t$. The constraints (17c) can be cast as the mixed-integer set

$$R_p := \left\{ (u_p, \mathbf{r}, \mathbf{z}) \in \mathbb{R} \times \mathbb{R}_+^N \times \{0, 1\}^N : u_p + r_i \geq h_{i,p}(1 - z_i), i \in [N] \right\}.$$

In fact, a similar mixed-integer set has been studied in the context of robust 0–1 programming by Atamtürk [2]:

$$R^+ := \left\{ (u, \mathbf{r}, \mathbf{z}) \in \mathbb{R}_+ \times \mathbb{R}_+^N \times \{0, 1\}^N : u + r_i \geq h_i(1 - z_i), i \in [N] \right\},$$

where h_1, \dots, h_N are assumed to all be positive. The main difference between R_p and R^+ is that u_p is unrestricted in R_p but is non-negative in R^+ , and some $h_{i,p}$ may be negative or zero in R_p , but all h_i are positive in R^+ . We refer to the set R^+ as the *robust 0–1 set*, since it originates from the robust counterpart of a 0–1 program whose objective vector belongs to a “budget uncertainty set”. In Atamtürk [2], the binary variables in R^+ correspond to the original decision variables of the 0–1 program.

The fact that R_p is so similar to R^+ may seem surprising at first, yet is less so if we consider the conditional value-at-risk (CVaR) interpretation for the Wasserstein robust chance constraint. More precisely, Xie [32, Corollary 1] gives the following alternate version of (3):

$$\mathcal{X}_{\text{DR}}(\mathcal{S}) = \left\{ \mathbf{x} \in \mathcal{X} : \frac{\theta}{\epsilon} + \text{CVaR}_{1-\epsilon}(-\text{dist}(\boldsymbol{\xi}, \mathcal{S}(\mathbf{x})); \mathbb{P}_N) \leq 0 \right\} \tag{18a}$$

$$= \left\{ \mathbf{x} \in \mathcal{X} : \frac{\theta}{\epsilon} + \max_{y \in B} \frac{1}{\epsilon N} \sum_{i \in [N]} (-\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x}))) y_i \leq 0 \right\} \tag{18b}$$

$$\text{where } \text{CVaR}_{1-\epsilon}(v(\boldsymbol{\xi}); \mathbb{P}_N) := \min_t \left\{ t + \frac{1}{\epsilon N} \sum_{i \in [N]} \max\{0, v(\boldsymbol{\xi}_i) - t\} \right\}, \tag{18c}$$

$$B := \left\{ \mathbf{y} : \frac{1}{N} \sum_{i \in [N]} y_i = \epsilon, \mathbf{0} \leq \mathbf{y} \leq \mathbf{1} \right\}. \tag{18d}$$

Note that (18b) comes from the dual interpretation of the CVaR, see e.g., [28, (3)]. In particular, B has exactly the same structure as the budget uncertainty set studied in Atamtürk [2].

We now further exploit the CVaR interpretation (18a) of (3) so that we can cast R_p exactly in the same form as R^+ . Specifically, we show that $u_p \geq 0$ is a valid inequality for (17), which will then allow us to eliminate constraints with negative $h_{i,p}$. Consequently, this allows us to directly apply the valid inequalities derived for R^+ from Atamtürk [2] to R_p . We first provide a bound on the value of the t -variable in (3) (which then translates to a bound on the t -variables for all subsequent formulations).

Lemma 1 *Let $k := \lfloor \epsilon N \rfloor$ and fix any $\mathbf{x} \in \mathcal{X}_{\text{DR}}(\mathcal{S})$. Then there exists (\mathbf{r}, t) such that t is equal to the $(k + 1)$ -th smallest value amongst $\{\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x}))\}_{i \in [N]}$ and the constraints of (3) are satisfied.*

Proof Recall that the CVaR of a random variable $v(\boldsymbol{\xi})$ from (18c), where $\boldsymbol{\xi} \sim \mathbb{P}_N$, has two equivalent primal and dual optimization representations:

$$\text{CVaR}_{1-\epsilon}(v(\boldsymbol{\xi}); \mathbb{P}_N) = \max_{\mathbf{y} \in B} \frac{1}{\epsilon N} \sum_{i \in [N]} v(\boldsymbol{\xi}_i) y_i \tag{19a}$$

$$= \min_{\mathbf{r}, t'} \left\{ t' + \frac{1}{\epsilon N} \sum_{i \in [N]} r_i : r_i \geq v(\boldsymbol{\xi}_i) - t', i \in [N], \mathbf{r} \geq \mathbf{0} \right\}, \tag{19b}$$

where B is defined in (18d). Without loss of generality, assume that we have an ordering $v(\boldsymbol{\xi}_1) \geq \dots \geq v(\boldsymbol{\xi}_N)$. It is easy to check that a primal-dual optimal pair for (19) is given by

$$y_i = \begin{cases} 1, & i = 1, \dots, k \\ \epsilon N - k, & i = k + 1 \\ 0, & i > k + 1 \end{cases}$$

$$t' = v(\boldsymbol{\xi}_{k+1})$$

$$r_i = \max \{0, v(\boldsymbol{\xi}_i) - t'\},$$

as $\mathbf{y} \in B$ and $\mathbf{r} \geq \mathbf{0}$, and

$$\frac{1}{\epsilon N} \sum_{i \in [N]} v(\boldsymbol{\xi}_i) y_i = t' + \frac{1}{\epsilon N} \sum_{i \in [N]} r_i.$$

Now take $v(\boldsymbol{\xi}_i) = -\text{dist}(\boldsymbol{\xi}_i, \mathcal{S}(\mathbf{x}))$ for each $i \in [N]$, and recall that we assume $v(\boldsymbol{\xi}_1) = -\text{dist}(\boldsymbol{\xi}_1, \mathcal{S}(\mathbf{x})) \geq \dots \geq v(\boldsymbol{\xi}_N) = -\text{dist}(\boldsymbol{\xi}_N, \mathcal{S}(\mathbf{x}))$. Let (\mathbf{r}, t') be the

optimal solution to the CVaR formulation (19b) specified above with $t' = v(\xi_{k+1}) = -\text{dist}(\xi_{k+1}, \mathcal{S}(\mathbf{x}))$, $r_i = \max\{0, v(\xi_i) - t'\}$ for all $i \in [N]$. Since $\mathbf{x} \in \mathcal{X}_{\text{DR}}(\mathcal{S})$, we know from (18a) that

$$\frac{\theta}{\epsilon} + \text{CVaR}_{1-\epsilon}(-\text{dist}(\xi, \mathcal{S}(\mathbf{x})); \mathbb{P}_N) \leq 0 \implies \theta + \frac{1}{N} \sum_{i \in [N]} r_i \leq -\epsilon t'.$$

Now take $t = -t' = \text{dist}(\xi_{k+1}, \mathcal{S}(\mathbf{x})) \geq 0$, and notice that $r_i = \max\{0, v(\xi_i) - t'\} = \max\{0, t - \text{dist}(\xi_i, \mathcal{S}(\mathbf{x}))\}$ for all $i \in [N]$, so the constraints in (3) are satisfied. \square

The main result for this section is to show that $u_p \geq 0$ is valid for (17) for each $p \in [P]$.

Proposition 1 *Suppose that $\theta > 0$. Consider an arbitrary $\mathbf{x} \in \mathcal{X}_{\text{DR}}(\mathcal{S})$. There exists $(\mathbf{r}, t, \mathbf{z})$ such that $(\mathbf{x}, \mathbf{r}, t, \mathbf{z})$ satisfies (17b)–(17c) and that for every $p \in [P]$,*

$$u_p = \frac{-q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} - t \geq 0.$$

Proof For convenience, for each $p \in [P]$, denote $g_{i,p}(\mathbf{x}) := \frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*}$ and $g_p^*(\mathbf{x}) := \frac{-q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*}$. With these definitions, the distance function from (4) is

$$\text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) = \max \left\{ 0, \min_{p \in [P]} g_{i,p}(\mathbf{x}) \right\}.$$

Without loss of generality, assume that $\text{dist}(\xi_1, \mathcal{S}(\mathbf{x})) \leq \dots \leq \text{dist}(\xi_N, \mathcal{S}(\mathbf{x}))$, and denote $d^*(\mathbf{x}) := \text{dist}(\xi_{k+1}, \mathcal{S}(\mathbf{x}))$ to be the $(k + 1)$ -smallest distance value.

By Lemma 1, take (\mathbf{r}, t) to be a solution that satisfies $t = d^*(\mathbf{x})$, $r_i = \max\{0, t - \text{dist}(\xi_i, \mathcal{S}(\mathbf{x}))\}$, and set $z_i = 1$ when $\min_{p \in [P]} g_{i,p}(\mathbf{x}) < 0$, otherwise $z_i = 0$ for each $i \in [N]$. It is straightforward to check that $(\mathbf{x}, \mathbf{r}, t, \mathbf{z})$ satisfies (17b), so we focus on (17c). If $z_i = 0$, then since $\text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) \geq t - r_i$ and $\text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) = \min_{p \in [P]} g_{i,p}(\mathbf{x})$, (17c) holds. If $z_i = 1$, then $\text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) = 0$, and $r_i = t$, so $t - r_i = 0$, hence (17c) reduces to (15a).

It remains to show that for every $p \in [P]$, $0 \leq u_p = g_p^*(x) - t = g_p^*(x) - d^*(x)$, i.e., $d^*(\mathbf{x}) \leq g_p^*(\mathbf{x})$ holds. Note that by definition of $-q_p$, g_p^* is the $(k + 1)$ -th smallest value amongst $\{g_{i,p}(\mathbf{x})\}_{i \in [N]}$. Focusing on the definition of $d^*(\mathbf{x}) := \text{dist}(\xi_{k+1}, \mathcal{S}(\mathbf{x}))$, we consider two cases.

If $\min_{p \in [P]} g_{k+1,p}(\mathbf{x}) \leq 0$, then $d^*(\mathbf{x}) = t = 0$. But, then we cannot have $\epsilon t \geq \theta + \frac{1}{N} \sum_{i \in [N]} r_i$, since $\theta > 0$ and $\mathbf{r} \geq \mathbf{0}$. Thus, we cannot have $\min_{p \in [P]} g_{k+1,p}(\mathbf{x}) \leq 0$.

Now consider $\min_{p \in [P]} g_{k+1,p}(\mathbf{x}) > 0$. Then, for any $i \geq k + 1$ and $p \in [P]$, we have

$$0 < \min_{p \in [P]} g_{k+1,p}(\mathbf{x}) = d^*(\mathbf{x}) \leq \text{dist}(\xi_i, \mathcal{S}(\mathbf{x})) = \min_{p' \in [P]} g_{i,p'}(\mathbf{x}) \leq g_{i,p}(\mathbf{x}),$$

where the second inequality follows from the fact that $i \geq k + 1$ and the assumption that $\text{dist}(\xi_1, \mathcal{S}(\mathbf{x})) \leq \dots \leq \text{dist}(\xi_N, \mathcal{S}(\mathbf{x}))$, and the last equation follows from the fact that $0 < \text{dist}(\xi_i, \mathcal{S}(\mathbf{x}))$. Based on this relation, then there are at least $N - k$ indices i such that $g_{i,p}(\mathbf{x}) \geq d^*(\mathbf{x})$, and thus we must have $g_p^*(\mathbf{x}) \geq d^*(\mathbf{x})$. This completes the proof. \square

We can now explicitly impose the constraint $u_p \geq 0$ into the formulation for (DR-CCP). When we do this, the constraint $u_p + r_i \geq h_{i,p}(1 - z_i)$ corresponding to any indices i, p such that $h_{i,p} \leq 0$ becomes redundant since the left-hand side of this constraint is non-negative, and its right-hand side is non-positive. Based on this, in our improved formulation, we define the following index sets:

$$[N]_p := \left\{ i \in [N] : -\mathbf{b}_p^\top \xi_i > q_p \right\}, \quad p \in [P].$$

Then, $h_{i,p} > 0$ if and only if $i \in [N]_p$. Our proposed formulation is as follows:

$$\min_{\mathbf{z}, \mathbf{r}, t, \mathbf{x}} \quad \mathbf{c}^\top \mathbf{x} \tag{20a}$$

$$\text{s.t. } (\mathbf{z}, \mathbf{r}, t, \mathbf{x}) \text{ satisfies (5b)–(5d) and (8c),} \tag{20b}$$

$$\frac{\mathbf{b}_p^\top \xi_i + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} + \frac{-\mathbf{b}_p^\top \xi_i - q_p}{\|\mathbf{b}_p\|_*} z_i \geq t - r_i, \quad i \in [N]_p, p \in [P], \tag{20c}$$

$$\frac{-q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} \geq t, \quad p \in [P]. \tag{20d}$$

Theorem 3 *Formulation (20) is an exact reformulation of (DR-CCP) where the safety set is given by (4).*

Proof The correctness of (20) follows from the above discussion. \square

Remark 2 The size of each index set $[N]_p$ is at most $k = \lfloor \epsilon N \rfloor$, so (20) reduces the number of constraints of (17) by at least $((1 - \epsilon)N - 1)P$. This is particularly significant when ϵ is small (e.g., 0.1) and N, P are large. \square

Note that similar bounding and scenario elimination strategies using the VaR interpretation are shown to be effective in multivariate CVaR-constrained optimization and risk-averse Markov decision processes [18,21,26,29].

Importantly, via the constraints (20c) and (20d), we can re-define the mixed-integer set R_p to have the exact same structure as R^+ :

$$R_p := \left\{ (u_p, \mathbf{r}, \mathbf{z}) \in \mathbb{R}_+ \times \mathbb{R}_+^N \times \{0, 1\}^N : u_p + r_i \geq h_{i,p}(1 - z_i), i \in [N]_p \right\},$$

where it is understood that $u_p = \frac{-q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} - t$. Atamtürk [2, Theorem 1] proposes a class of valid inequalities for R^+ , which we adapt to obtain valid inequalities for

R_p . To do this, we let N_p be the size of each $[N]_p$ and define an ordering $[N]_p = \{(j, p) \in \mathbb{N} : j \in [N_p]\}$ as follows:

$$h_{(N_p,p),p} \geq \dots \geq h_{(1,p),p}.$$

Proposition 2 For any $p \in [P]$ and $J = \{j_1, \dots, j_m\}$ satisfying $m \geq 1, N_p \geq j_1 \geq \dots \geq j_m \geq 1$, the following inequality is valid for (20):

$$\frac{-q_p + d_p - \mathbf{a}_p^\top \mathbf{x}}{\|\mathbf{b}_p\|_*} - t + \sum_{i \in [m]} r_{(j_i,p)} \geq \sum_{i \in [m]} (h_{(j_i,p),p} - h_{(j_{i+1},p),p})(1 - z_{(j_i,p)}) \tag{21}$$

where $h_{(j_{m+1},p),p} := 0$.

Proof This follows immediately from Atamtürk [2, Theorem 1]. □

We refer to the inequalities (21) as the *path inequalities*.

Remark 3 Atamtürk [2] gives an $O(N_p^2) = O(\lfloor \epsilon N \rfloor^2)$ -time separation algorithm for (21), which is based on finding a shortest path in an acyclic graph. Atamtürk [2, Theorems 1 and 2] also proves that inequalities (21) are sufficient to describe the convex hull of R^+ and are facet-defining. □

5 Computational study

In this section, we assess the numerical performance of our improved formulation (20) of (DR-CCP) and valid inequalities from Sects. 3 and 4.

All experiments are conducted on an Intel Core i5 3GHz processor with 6 cores and 32GB memory. Each experiment was in single-core mode, and five experiments were run in parallel. For each model, we set the CPLEX time limit to be 3600 s.

CPLEX 12.9 is used as the MIP solver. Valid inequalities from Sects. 3 and 4 are separated and added via the CPLEX user-cut callback feature. Since using a user-cut callback function is known to affect various internal CPLEX dynamics (such as dynamic search, aggressiveness of CPLEX presolve and cut generation procedures, etc.), in order to do a fair comparison, we include an empty user-cut callback function (that does not separate any user cuts) whenever we test a formulation which does not employ cuts, e.g., basic formulation from the literature, i.e., from Chen et al. [8]. We have also conducted tests under the default CPLEX settings without the empty user cut callback to confirm that our overall conclusions do not change under the default settings. Our preliminary tests indicated that separating a large number of inequalities throughout the branch-and-cut tree slows down the search process, so we separate our inequalities only at the root node.

The maximum value of θ at which (DR-CCP) is feasible can be computed by solving a variant of (5), that is, with the same set of constraints (5b)–(5e) but treating θ as a variable to be maximized. We first describe our test instances in Sect. 5.1 and then discuss the performance of our proposed approaches in Sect. 5.2.

5.1 Test instances

We consider the distributionally robust chance-constrained formulation of a transportation problem from Chen et al. [8]. This is the problem of transshipping a single good from a set of factories $[F]$ to a set of distribution centers $[D]$ to meet their demands while minimizing the transportation cost. Each factory $f \in [F]$ has an individual production capacity m_f , each distribution center $d \in [D]$ faces a random demand ξ_d from the end customers, and transshipping one unit of the good from factory f to distribution center incurs a cost c_{fd} . Given N samples $\{\xi_i = (\xi_{id})_{d \in [D]} : i \in [N]\}$ of ξ , this problem is given by

$$\min \mathbf{c}^\top \mathbf{x} \tag{22a}$$

$$\text{s.t. } \mathbb{P} \left[\sum_{f \in [F]} x_{fd} \geq \xi_d, \quad \forall d \in [D] \right] \geq 1 - \epsilon, \quad \mathbb{P} \in \mathcal{F}(\theta), \tag{22b}$$

$$\sum_{d \in [D]} x_{fd} \leq m_f, \quad f \in [F], \tag{22c}$$

$$x_{fd} \geq 0, \quad f \in [F], \quad d \in [D]. \tag{22d}$$

Here, (22b) is a joint chance constraint with right-hand side uncertainty, so (22) can be reformulated as in (5).

We use the same random instance generation scheme from Chen et al [8]. We generate instances with $F \in \{5, 10\}$ factories and $D \in \{50, 100\}$ distribution centers whose locations are chosen uniformly at random from the Euclidean plane $[0, 10]^2$. We set the transportation cost c_{fd} to the Euclidean distance between factory $f \in [F]$ and distribution center $d \in [D]$. We obtain the scenarios by sampling for the demand vector ξ from a uniform distribution supported on $[0.8\mu, 1.2\mu]$, where the expected demand μ_d of any distribution center $d \in [D]$ is chosen uniformly at random from $[0, 10]$. The capacity m_f of each factory $f \in [F]$ is drawn uniformly from $[0, 1]$ at first, but the capacities are scaled later so that the total capacity $\sum_{f \in [F]} m_f$ equals $\frac{3}{2} \max_{i \in [N]} \{\sum_{d \in [D]} \xi_{id}\}$. For each instance, we test ten different values $\theta_1 < \dots < \theta_{10}$ for the Wasserstein radius. As in Chen et al.[8], we set $\theta_1 = 0.001$. For the other values, we compute the maximum value θ_{\max} of θ such that (5) is feasible and set $\theta_j = \frac{j-1}{10} \theta_{\max}$ for $j = 2, \dots, 10$. We have empirically found that the value θ_{\max} is between 0.1 and 0.35 for our instances, so θ_2 is between 0.01 and 0.035 and thus greater than $\theta_1 = 0.001$. We fix the risk tolerance ϵ to be 0.1. Lastly, we select M to be

$$M := \max \left\{ \sum_{f \in [F]} m_f - \min_{i \in [N], d \in [D]} \{\xi_{id}\}, \max_{i \in [N], d \in [D]} \{\xi_{id}\} \right\}, \tag{23}$$

which is sufficiently large (see Remark 1). For each problem parameter combination, we generate 10 random instances and report the average statistics.

As noted in Chen et al. [8], there is no need to specify which norm $\|\cdot\|$ to use in (5) and (8) when reformulating (22) because the random right-hand side inside (22b) contains a single random variable with coefficient 1 so that all $\|\mathbf{b}_p\|_*$ in (5) and (8) equal 1 in the instances generated.

5.2 Performance analysis

In this section, we summarize our experiments with radius $\theta = \theta_1, \dots, \theta_{10}$ and the number of samples $N = 100, 1000, 3000$. We compare the following five formulations:

- Basic: the basic formulation (5) given by Chen et al. [8] with the big- M value computed as in Remark 1,
- Improved: the improved formulation (20),
- Mixing: the improved formulation (20) with the mixing inequalities (16),
- Path: the improved formulation (20) with the path inequalities (21), and
- Mixing+Path: the improved formulation (20) with both mixing (16) and path inequalities (21).

In Tables 1 and 3, the following statistics are reported:

Time(Gap): the average solution time (in seconds measured externally from CPLEX by C++) of the instances that were solved to optimality, and, in parentheses, the average of the final optimality gap of the instances that were not solved to optimality within the CPLEX time limit. The optimality gap is computed as $(UB - LB)/LB * 100$ where UB and LB respectively are the objective values of the best feasible solution and the best lower bound value at the corresponding time of the solution process. A ‘*’ in a Time or Gap entry indicates that either no instance was solved to optimality or all instances were solved to optimality within the CPLEX time limit so that there were no instances for measuring the corresponding statistic.

We also report in this column the number of instances solved to optimality within the CPLEX time limit, s , and the number of instances for which a feasible solution was found f . Since for most cases, we observed that $s = f = 10$, we add $[s/f]$ in front of the Time(Gap) statistic only when $s < 10$ or $f < 10$.

A ‘n/a’ entry for the Time(Gap) statistic denotes when no feasible solution was found in any of the 10 instances.

Cuts: the average number of cuts added for Mixing, Path, and Mixing+Path. For Mixing+Path, the ‘Cuts’ are broken down into the number of mixing inequalities and the number of path inequalities added respectively.

In Tables 2 and 4, the following statistics are reported:

R.time: the average time spent (in seconds measured externally from CPLEX by C++) at the root node of the branch-and-bound tree

over all instances. A ‘n/a’ entry indicates that no feasible solution was found in any of the 10 instances within the CPLEX time limit.

R.gap: the final optimality gap at the root node of the branch-and-bound tree. A ‘n/a’ entry indicates that no solution was found in any of the 10 instances within the CPLEX time limit.

The results in all tables highlight that when the radius θ is small, the resulting problems are much harder to solve. This was also reported in Chen et al. [8].

When $N = 100$, $F = 5$, $D = 50$, and $\theta = \theta_1$, Table 1 shows that `Basic` does not finish within the CPLEX time limit of 3600s for any of the 10 randomly generated instances and terminates with a 1.16% optimality gap, on average. In contrast, `Improved` solves all instances to optimality in under 5s on average. In fact, `Improved` is so effective that it does not leave much room for improvement for the additional valid inequalities in `Mixing`, `Path` and `Mixing+Path`, and the separation of the valid inequalities results in a slight increase in the solution time in most cases. Nevertheless, the latter three formulations solve all instances to optimality in under 9s on average. When $\theta \geq \theta_2$, `Basic` solves all instances to optimality in under 27s on average, but all other formulations solve in under 0.06s on average. In “Appendix 1”, we provide supplementary results that show that the mixing and path inequalities are very effective when added to `Basic`, but `Improved` without any valid inequalities performs better than `Basic` with these inequalities.

To test the scalability of our proposed approaches with respect to the number of scenarios, in Table 1, we also report the performance of the five formulations on instances with $N = 1000$ and $N = 3000$, respectively for $F = 5$, $D = 50$. In the out-of-sample tests for these instances reported in [8], for $N = 1000$, the authors state the difficulty of solving the problem to proven optimality and report their results with a not necessarily optimal solution obtained after a couple of minutes of computing. Our results in Table 1 show that our proposed formulation provides provably optimal solutions within less than a minute in all instances but those with θ_1 . Furthermore, as indicated in Table 1, we can scale up the number of scenarios even further. In the experiments with $N = 3000$, we observe that even for the largest value of θ_{10} , `Basic` was unable to solve any of the ten instances within the CPLEX time limit, while all of the new formulations `Improved`, `Mixing`, `Path` and `Mixing+Path` solved all instances to optimality for $\theta \geq \theta_3$ with an average time of at most 20s. For $\theta \leq \theta_2$, our new formulations did not manage to solve any instances to optimality, but did reduce the average optimality gap to at most 0.8%. In contrast, `Basic` failed to solve all instances for $\theta \leq \theta_2$, and in fact did not even find a feasible solution within the time limit for many instances. For the instances where a feasible solution was found, the gap remained quite large at $\approx 70\%$ at termination.

We also test the scalability of various methods with respect to the number of original decision variables in the original problem. To this end, we let $F = 10$, $D = 100$ and report our results in Table 3 for $N = 100, 1000, 3000$. (Note that an increase in D implies an increase in the dimension of the random data, which may in turn

Table 1 Results for $F = 5, D = 50, \epsilon = 0.1$

N	θ	Basic		Improved		Mixing		Path		Mixing+Path	
		Time(Gap)	Cuts	Time(Gap)	Cuts	Time(Gap)	Cuts	Time(Gap)	Cuts	Time(Gap)	Cuts
100	θ_1	[0/10] *(1.16)	4.29(*)	6.64(*)	54.7	7.31(*)	287.6	8.40(*)	41.7/274.6		
	θ_2	26.58(*)	0.04(*)	0.05(*)	0.3	0.06(*)	88.2	0.06(*)	0.3/88.2		
	θ_3	4.27(*)	0.04(*)	0.04(*)	0.0	0.04(*)	73.8	0.05(*)	0.0/73.8		
	θ_4	3.36(*)	0.03(*)	0.03(*)	0.0	0.03(*)	44.4	0.03(*)	0.0/44.4		
	θ_5	1.34(*)	0.03(*)	0.03(*)	0.0	0.03(*)	42.4	0.03(*)	0.0/42.4		
	θ_6	1.17(*)	0.03(*)	0.03(*)	0.0	0.03(*)	43.1	0.03(*)	0.0/43.1		
	θ_7	1.04(*)	0.03(*)	0.03(*)	0.0	0.03(*)	40.3	0.03(*)	0.0/40.3		
	θ_8	0.94(*)	0.03(*)	0.03(*)	0.0	0.03(*)	40.2	0.03(*)	0.0/40.2		
	θ_9	0.86(*)	0.03(*)	0.03(*)	0.0	0.03(*)	36.4	0.03(*)	0.0/36.4		
	θ_{10}	0.84(*)	0.02(*)	0.03(*)	0.0	0.03(*)	36.0	0.03(*)	0.0/36.0		
1000	θ_1	[0/6] *(97.90)	[0/10] *(0.59)	[0/10] *(0.52)	532.8	[0/10] *(0.44)	2472.5	[0/10] *(0.44)	381.0/2377.9		
	θ_2	[0/10] *(66.04)	36.84(*)	36.68(*)	0.0	40.41(*)	114.3	40.61(*)	0.0/114.3		
	θ_3	[0/10] *(31.10)	9.15(*)	9.33(*)	0.0	9.23(*)	82.9	9.61(*)	0.0/82.9		
	θ_4	[0/10] *(2.11)	7.73(*)	7.60(*)	0.0	7.74(*)	61.6	7.71(*)	0.0/61.6		
	θ_5	287.92(*)	6.12(*)	6.16(*)	0.0	6.20(*)	58.5	6.19(*)	0.0/58.5		
	θ_6	130.38(*)	5.83(*)	5.83(*)	0.0	5.95(*)	49.0	5.94(*)	0.0/49.0		
	θ_7	117.77(*)	5.75(*)	5.72(*)	0.0	5.77(*)	48.9	5.75(*)	0.0/48.9		
	θ_8	99.95(*)	5.58(*)	5.52(*)	0.0	5.65(*)	49.0	5.64(*)	0.0/49.0		
	θ_9	90.82(*)	5.27(*)	5.27(*)	0.0	5.34(*)	49.0	5.39(*)	0.0/49.0		
	θ_{10}	78.88(*)	4.95(*)	4.97(*)	0.0	5.04(*)	49.1	5.02(*)	0.0/49.1		

Table 1 continued

N	θ	Basic		Improved		Mixing		Path		Mixing+Path	
		Time(Gap)	Cuts	Time(Gap)	Cuts	Time(Gap)	Cuts	Time(Gap)	Cuts	Time(Gap)	Cuts
3000	θ_1	[0/0] n/a		[0/10] *(0.78)	2358.2	[0/10] *(0.52)		[0/10] *(0.62)		[0/10] *(0.48)	1470.3/4228.1
	θ_2	[0/5] *(69.56)		[0/10] *(0.49)	0.0	[0/10] *(0.49)		[0/10] *(0.41)		[0/10] *(0.41)	0.0/6102.2
	θ_3	[0/4] *(48.65)		17.89(*)	0.0	17.62(*)	0.0	18.94(*)		18.29(*)	0.0/200.8
	θ_4	[0/4] *(15.01)		13.74(*)	0.0	13.22(*)	0.0	14.03(*)		13.94(*)	0.0/94.1
	θ_5	[0/10] *(1.11)		12.75(*)	0.0	12.60(*)	0.0	13.65(*)		13.55(*)	0.0/88.3
	θ_6	[0/10] *(0.58)		12.29(*)	0.0	12.21(*)	0.0	12.83(*)		12.68(*)	0.0/80.2
	θ_7	[0/10] *(0.48)		12.28(*)	0.0	12.01(*)	0.0	12.93(*)		12.95(*)	0.0/56.7
	θ_8	[0/10] *(0.40)		12.48(*)	0.0	12.20(*)	0.0	12.78(*)		12.65(*)	0.0/52.9
	θ_9	[0/10] *(0.37)		11.96(*)	0.0	11.81(*)	0.0	12.36(*)		12.52(*)	0.0/49.4
	θ_{10}	[0/10] *(0.28)		8.04(*)	0.0	8.19(*)	0.0	8.86(*)		8.94(*)	0.0/49.4

Table 2 Results at root node for $F = 5, D = 50, \epsilon = 0.1$

N	θ	Basic		Improved		Mixing		Path		Mixing+Path	
		R.time	R.gap	R.time	R.gap	R.time	R.gap	R.time	R.gap	R.time	R.gap
100	θ_1	5.67	90.53	0.19	0.34	0.21	0.35	0.50	0.31	0.47	0.31
	θ_2	11.22	65.19	0.04	0.00	0.05	0.00	0.06	0.00	0.06	0.00
	θ_3	2.98	38.94	0.04	0.00	0.04	0.00	0.04	0.00	0.05	0.00
	θ_4	2.66	13.35	0.03	0.00	0.03	0.00	0.03	0.00	0.03	0.00
	θ_5	1.32	0.04	0.03	0.00	0.03	0.00	0.03	0.00	0.03	0.00
	θ_6	1.17	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.03	0.00
	θ_7	1.04	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.03	0.00
	θ_8	0.94	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.03	0.00
	θ_9	0.86	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.03	0.00
	θ_{10}	0.84	0.00	0.02	0.00	0.03	0.00	0.03	0.00	0.03	0.00
	1000	θ_1	2334.91	99.45	338.01	0.63	614.52	0.54	1968.43	0.45	2000.91
θ_2		275.76	73.64	37.57	0.01	36.76	0.01	40.65	0.00	40.48	0.00
θ_3		223.92	47.13	9.22	0.00	9.30	0.00	9.29	0.00	9.59	0.00
θ_4		289.21	16.30	7.70	0.00	7.71	0.00	7.77	0.00	7.77	0.00
θ_5		103.41	0.51	6.10	0.00	6.09	0.00	6.19	0.00	6.18	0.00
θ_6		89.56	0.39	5.81	0.01	5.83	0.01	5.92	0.01	5.86	0.01
θ_7		78.75	0.41	5.76	0.00	5.71	0.00	5.74	0.00	5.80	0.00
θ_8		67.40	0.40	5.53	0.00	5.55	0.00	5.62	0.00	5.65	0.00
θ_9		63.73	0.38	5.28	0.00	5.26	0.00	5.35	0.00	5.36	0.00
θ_{10}		58.08	0.36	4.93	0.00	4.91	0.00	5.04	0.00	5.03	0.00

Table 2 continued

N	θ	Basic		Improved		Mixing		Path		Mixing+Path	
		R.time	R.gap	R.time	R.gap	R.time	R.gap	R.time	R.gap	R.time	R.gap
3000	θ_1	n/a	n/a	72.08	0.80	1931.99	0.52	3598.81	0.62	3601.05	0.48
	θ_2	3144.09	70.41	134.46	0.55	109.72	0.55	3600.10	0.41	3600.22	0.41
	θ_3	2952.26	51.31	17.89	0.01	17.62	0.01	18.94	0.01	18.29	0.01
	θ_4	2684.77	15.72	13.74	0.01	13.22	0.01	14.03	0.01	13.94	0.01
	θ_5	3181.43	1.14	12.75	0.00	12.60	0.00	13.65	0.00	13.55	0.00
	θ_6	3176.11	0.63	12.29	0.00	12.21	0.00	12.83	0.00	12.68	0.00
	θ_7	2958.81	0.55	12.28	0.01	12.01	0.01	12.93	0.01	12.95	0.01
	θ_8	2876.49	0.47	12.48	0.01	12.20	0.01	12.78	0.01	12.65	0.01
	θ_9	2781.77	0.45	11.96	0.01	11.81	0.01	12.36	0.01	12.52	0.01
	θ_{10}	2439.69	0.41	8.04	0.01	8.19	0.01	8.86	0.01	8.94	0.01

Table 3 Results for $F = 10, D = 100, \epsilon = 0.1$

N	θ	Basic		Improved		Mixing		Path		Mixing+Path	
		Time(Gap)	Cuts	Time(Gap)	Cuts	Time(Gap)	Cuts	Time(Gap)	Cuts	Time(Gap)	Cuts
100	θ_1	[0/10] *(3.01)	10.46(*)	19.19(*)	55.7	29.36(*)	228.2	35.45(*)	48.0/236.2		
	θ_2	37.49(*)	0.11(*)	0.11(*)	0.0	0.12(*)	171.3	0.13(*)	0.0/171.3		
	θ_3	18.45(*)	0.10(*)	0.10(*)	0.0	0.11(*)	146.7	0.12(*)	0.0/146.7		
	θ_4	16.91(*)	0.09(*)	0.09(*)	0.0	0.10(*)	123.2	0.10(*)	0.0/123.2		
	θ_5	9.17(*)	0.09(*)	0.09(*)	0.0	0.09(*)	119.0	0.09(*)	0.0/119.0		
	θ_6	5.71(*)	0.06(*)	0.06(*)	0.0	0.07(*)	84.7	0.07(*)	0.0/84.7		
	θ_7	4.74(*)	0.06(*)	0.06(*)	0.0	0.07(*)	86.3	0.07(*)	0.0/86.3		
	θ_8	3.78(*)	0.06(*)	0.06(*)	0.0	0.07(*)	74.5	0.07(*)	0.0/74.5		
	θ_9	3.20(*)	0.06(*)	0.06(*)	0.0	0.06(*)	69.9	0.06(*)	0.0/69.9		
	θ_{10}	2.89(*)	0.05(*)	0.06(*)	0.0	0.06(*)	54.8	0.06(*)	0.0/54.8		
1000	θ_1	[0/1] *(98.73)	[0/10] *(0.45)	[0/10] *(0.42)	382.3	[0/10] *(0.39)	2150.7	[0/10] *(0.39)	249.6/2032.4		
	θ_2	[0/7] *(67.05)	27.02(*)	26.05(*)	0.0	37.06(*)	399.8	36.97(*)	0.0/399.8		
	θ_3	[0/4] *(34.22)	14.44(*)	14.73(*)	0.0	15.55(*)	231.1	14.70(*)	0.0/231.1		
	θ_4	[0/5] *(10.48)	10.18(*)	10.41(*)	0.0	12.13(*)	186.3	12.07(*)	0.0/186.3		
	θ_5	[4/10] 2599.60(0.45)	10.37(*)	10.68(*)	0.0	11.66(*)	186.1	11.31(*)	0.0/186.1		
	θ_6	968.46(*)	3.14(*)	3.25(*)	0.0	3.29(*)	96.3	3.28(*)	0.0/96.3		
	θ_7	853.68(*)	2.91(*)	3.06(*)	0.0	3.48(*)	96.4	3.52(*)	0.0/96.4		
	θ_8	647.80(*)	3.08(*)	3.11(*)	0.0	3.12(*)	96.3	3.07(*)	0.0/96.3		
	θ_9	528.73(*)	2.78(*)	2.74(*)	0.0	2.81(*)	96.2	2.85(*)	0.0/96.2		
	θ_{10}	375.32(*)	2.58(*)	2.57(*)	0.0	2.60(*)	96.5	2.59(*)	0.0/96.5		

Table 3 continued

N	θ	Basic		Improved		Mixing		Path		Mixing+Path	
		Time(Gap)	Cuts	Time(Gap)	Cuts	Time(Gap)	Cuts	Time(Gap)	Cuts	Time(Gap)	Cuts
3000	θ_1	[0/0] n/a	1494.1	[0/10] *(0.59)	0.0	[0/10] *(0.48)	6003.2	[0/10] *(0.46)	6003.2	[0/10] *(0.44)	948.3/5356.1
	θ_2	[0/0] n/a	0.0	[0/10] *(0.49)	0.0	[0/10] *(0.49)	6117.0	[0/10] *(0.42)	6117.0	[0/10] *(0.42)	0.0/6029.7
	θ_3	[0/0] n/a	0.0	407.31(*)	0.0	336.00(*)	1422.3	541.10(*)	1422.3	642.69(*)	0.0/1422.3
	θ_4	[0/0] n/a	0.0	158.82(*)	0.0	158.30(*)	386.3	148.37(*)	386.3	159.23(*)	0.0/386.3
	θ_5	[0/0] n/a	0.0	64.38(*)	0.0	64.42(*)	293.8	99.60(*)	293.8	100.18(*)	0.0/293.8
	θ_6	[0/0] n/a	0.0	33.52(*)	0.0	33.99(*)	98.4	35.09(*)	98.4	35.16(*)	0.0/98.4
	θ_7	[0/0] n/a	0.0	33.33(*)	0.0	33.73(*)	98.6	35.30(*)	98.6	35.30(*)	0.0/98.6
	θ_8	[0/0] n/a	0.0	34.34(*)	0.0	34.81(*)	98.7	36.85(*)	98.7	37.01(*)	0.0/98.7
	θ_9	[0/0] n/a	0.0	35.38(*)	0.0	35.31(*)	98.6	36.69(*)	98.6	36.63(*)	0.0/98.6
	θ_{10}	[0/0] n/a	0.0	26.28(*)	0.0	25.79(*)	98.8	26.19(*)	98.8	25.76(*)	0.0/98.8

Table 4 Results at root node for $F = 10, D = 100, \epsilon = 0.1$

N	θ	Basic		Improved		Mixing		Path		Mixing+Path	
		R.time	R.gap	R.time	R.gap	R.time	R.gap	R.time	R.gap	R.time	R.gap
100	θ_1	29.20	90.99	0.38	0.42	0.35	0.44	0.46	0.43	0.50	0.43
	θ_2	10.39	68.59	0.11	0.00	0.11	0.00	0.12	0.00	0.13	0.00
	θ_3	14.11	43.85	0.10	0.00	0.10	0.00	0.11	0.00	0.12	0.00
	θ_4	14.50	21.52	0.09	0.00	0.09	0.00	0.10	0.00	0.10	0.00
	θ_5	8.97	1.12	0.09	0.00	0.09	0.00	0.09	0.00	0.09	0.00
	θ_6	5.66	0.01	0.06	0.00	0.06	0.00	0.07	0.00	0.07	0.00
	θ_7	4.76	0.01	0.06	0.00	0.06	0.00	0.07	0.00	0.07	0.00
	θ_8	3.78	0.00	0.06	0.00	0.06	0.00	0.07	0.00	0.07	0.00
	θ_9	3.20	0.00	0.06	0.00	0.06	0.00	0.06	0.00	0.06	0.00
	θ_{10}	2.89	0.01	0.05	0.00	0.06	0.00	0.06	0.00	0.06	0.00
	1000	θ_1	3542.86	99.48	29.81	0.50	35.95	0.45	148.13	0.41	157.22
θ_2		1185.36	72.76	27.02	0.01	26.05	0.01	37.06	0.00	36.97	0.00
θ_3		990.87	46.39	14.44	0.01	14.73	0.01	15.55	0.01	14.70	0.01
θ_4		1547.07	21.38	10.18	0.01	10.41	0.01	12.13	0.00	12.07	0.00
θ_5		570.64	0.84	10.37	0.00	10.68	0.00	11.66	0.00	11.31	0.00
θ_6		498.69	0.41	3.14	0.00	3.25	0.00	3.29	0.00	3.28	0.00
θ_7		453.85	0.43	2.91	0.00	3.06	0.00	3.48	0.00	3.52	0.00
θ_8		412.37	0.40	3.08	0.00	3.11	0.00	3.12	0.00	3.07	0.00
θ_9		366.56	0.36	2.78	0.00	2.74	0.00	2.81	0.00	2.85	0.00
θ_{10}		314.52	0.25	2.58	0.00	2.57	0.00	2.60	0.00	2.59	0.00

Table 4 continued

N	θ	Basic		Improved		Mixing		Path		Mixing+Path	
		R.time	R.gap	R.time	R.gap	R.time	R.gap	R.time	R.gap	R.time	R.gap
3000	θ_1	n/a	n/a	289.73	0.60	902.49	0.48	3599.15	0.46	3603.49	0.44
	θ_2	n/a	n/a	456.10	0.52	368.56	0.52	3601.21	0.42	3597.16	0.42
	θ_3	n/a	n/a	407.31	0.01	336.00	0.01	541.10	0.01	642.69	0.01
	θ_4	n/a	n/a	158.82	0.01	158.30	0.01	148.37	0.01	159.23	0.01
	θ_5	n/a	n/a	64.38	0.01	64.42	0.01	99.60	0.01	100.18	0.01
	θ_6	n/a	n/a	33.52	0.01	33.99	0.01	35.09	0.01	35.16	0.01
	θ_7	n/a	n/a	33.33	0.00	33.73	0.00	35.30	0.00	35.30	0.00
	θ_8	n/a	n/a	34.34	0.00	34.81	0.00	36.85	0.00	37.01	0.00
	θ_9	n/a	n/a	35.38	0.00	35.31	0.00	36.69	0.00	36.63	0.00
	θ_{10}	n/a	n/a	26.28	0.00	25.79	0.00	26.19	0.00	25.76	0.00

require an increase in sample size to ensure the same out-of-sample performance.) Comparing Tables 1 and 3, we see that not surprisingly, the problems with a larger number of original decision variables are harder to solve. Fewer instances can be solved to optimality with `Basic`. In fact, `Basic` cannot even obtain a feasible solution after an hour of computing for instances with $F = 10$, $D = 100$, $N = 3000$. In contrast, `Improved` solves all instances to optimality in less than a minute except for the ones with θ_1 when $N = 100, 1000$ and with $\theta \leq \theta_5$ in the case of $N = 3000$. Indeed, we observe a very slight increase in solution times for our proposed formulations in the case of $F = 10$, $D = 100$ in contrast to $F = 5$, $D = 50$, and all trends reported for $F = 5$, $D = 50$ remain the same in the case of $F = 10$, $D = 100$.

Tables 1 and 3 also give us insight into the marginal effect of each class of inequalities. We observe that mixing inequalities are only generated for $\theta = \theta_1$ (plus a total of 3 inequalities across all instances when $\theta = \theta_2$ and $N = 100$). When the radius of the Wasserstein ball is small, the nominal region $\mathcal{X}_{\text{SAA}}(\mathcal{S})$ is a better approximation for the distributionally robust region $\mathcal{X}_{\text{DR}}(\mathcal{S})$. Since mixing inequalities are valid for $\mathcal{X}_{\text{SAA}}(\mathcal{S})$, it is expected that they have stronger effects as for smaller radius θ , thus the observed behavior is not surprising. As mentioned above, when $N = 100$, the `Improved` formulation is already so effective that separating inequalities slightly increases solution times. However, when $N = 3000$, $F = 5$ and $D = 50$, from Table 1, we see that while `Improved` is still quite effective, and manages to reduce the gap to 0.78% for $\theta = \theta_1$, separating the valid inequalities reduces this further (0.52% for `Mixing`, 0.62% for `Path`, and 0.48% for `Mixing+Path`). We see a similar phenomenon for $\theta = \theta_2$, but only path inequalities are separated. Similar observations can be made for $N = 3000$, $F = 10$, $D = 100$.

Tables 2 and 4 provide further information on the performance of the formulations at the root node of the branch-and-bound tree. From Table 2, we see that the root gap of the new formulations `Improved`, `Mixing`, `Path` and `Mixing+Path` are at most 0.8% on average for the most difficult regime $\theta = \theta_1, \theta_2$ and $N = 3000$. This is significantly better than the root gap of `Basic`. In fact, in our experiments, we have observed that for these types of distributionally robust CCPs, the branch-and-bound process is very ineffective in terms of reducing the remaining gap, with only a small difference between root gap and final gap. This also highlights the importance of starting off with very strong formulations. In fact, we observed that for $\theta \geq \theta_2$, the gap threshold of 0.01% is achieved at the root node for our improved formulations.

6 Conclusion

This paper studies in detail the formulation for distributionally robust chance-constrained programs under Wasserstein ambiguity, focusing on the case of linear safety sets with right-hand side uncertainty. We reveal a hidden connection with nominal chance-constraints, and provide a class of valid inequalities which are exactly the mixing inequalities for the nominal chance constraint. We also adapt the quantile

strengthening technique to the distributionally robust setting, making one set of constraints (5e) in the original MIP formulation (5) big- M -free. We then exploit the CVaR interpretation of the distributionally robust constraint (18a) to provide an improved formulation (20) with significantly fewer constraints. Finally, we uncover a mixed-integer substructure which has been studied in the context of robust 0–1 programming Atamtürk [2], and provide second class of valid inequalities based on this connection.

Our computational results demonstrate the benefit of our improved formulation and valid inequalities. Solution times are drastically reduced and larger problems can now be solved in seconds, e.g., problems with thousands of scenarios (rather than hundreds).

Acknowledgements This paper is in memory of Shabbir Ahmed, whose fundamental contributions on mixing sets, chance-constrained programming and distributionally robust optimization we build upon. We thank the two referees and the AE for their suggestions that improved the exposition. This research is supported, in part, by ONR Grant N00014-19-1-2321, by the Institute for Basic Science (IBS-R029-C1), Award N660011824020 from the DARPA Lagrange Program and NSF Award 1740707.

Supplementary Numerical Results

In Table 5 we report the performance of the basic formulation (5) when combined with mixing (16) and path inequalities (21) for $F = 5$, $D = 50$, $\epsilon = 0.1$. These results highlight that mixing and path inequalities are indeed useful when applied to the basic formulation, without all the other enhancements we propose.

We see that for $N = 100$, `Basic+Mixing+Path` solves all instances (and with slightly quicker times on average) whereas `Basic` in Table 1 does not manage to solve any instances for θ_1 . For $N = 1000$, we again see an improvement in the number of instances solved when using the inequalities, but for larger $\theta_8, \theta_9, \theta_{10}$, times are slightly slower. We believe this is due to the extra time required to solve the larger LP relaxations as a result of adding cuts. For $N = 3000$, we now see that `Basic+Mixing+Path` is unable to find a feasible integer solution within 1 h for a larger number of instances than `Basic` (we again believe this is due to larger LP relaxations), but whenever it does, it often solves to optimality, which `Basic` never does.

In contrast to `Improved`, there are more cuts generated for `Basic` for all θ values. Nevertheless, as expected, the performance achieved by `Basic+Mixing+Path` is still worse than the performance of our improved formulation (20), with or without inequalities.

Table 5 Supplementary results for $F = 5$, $D = 50$, $\epsilon = 0.1$

N	θ	Basic+Mixing+Path			
		Time(Gap)	R.time	R.gap	Cuts
100	θ_1	36.64(*)	2.77	0.30	116.0/386.2
	θ_2	0.61(*)	0.61	0.00	65.3/148.7
	θ_3	0.64(*)	0.64	0.00	59.5/128.5
	θ_4	0.53(*)	0.53	0.00	57.7/118.2
	θ_5	0.56(*)	0.56	0.00	57.1/107.9
	θ_6	0.53(*)	0.53	0.00	55.5/101.0
	θ_7	0.43(*)	0.43	0.00	54.0/97.8
	θ_8	0.46(*)	0.46	0.00	53.2/96.9
	θ_9	0.51(*)	0.51	0.00	52.7/95.6
	θ_{10}	0.48(*)	0.48	0.00	51.7/92.4
1000	θ_1	[0/9] *(0.45)	2532.66	0.45	1047.7/3462.8
	θ_2	272.45(*)	272.45	0.00	320.0/1081.3
	θ_3	142.32(*)	142.32	0.00	220.2/620.2
	θ_4	125.95(*)	125.95	0.00	196.1/570.2
	θ_5	111.84(*)	111.84	0.00	186.4/521.6
	θ_6	106.86(*)	106.86	0.00	172.4/499.5
	θ_7	100.38(*)	100.38	0.00	162.2/461.5
	θ_8	103.55(*)	103.55	0.00	155.5/496.6
	θ_9	109.29(*)	109.29	0.00	144.7/464.5
	θ_{10}	107.14(*)	107.14	0.00	136.5/462.2
3000	θ_1	[0/0] n/a	n/a	n/a	n/a
	θ_2	[0/0] n/a	n/a	n/a	n/a
	θ_3	[0/0] n/a	n/a	n/a	n/a
	θ_4	[3/6] 3334.13(0.02)	3464.71	0.01	425.2/1091.0
	θ_5	[4/6] 3364.70(0.03)	3440.82	0.01	368.2/1120.7
	θ_6	[3/3] 3116.18(*)	3116.18	0.00	320.3/1188.0
	θ_7	[2/2] 3304.96(*)	3304.96	0.01	292.0/912.0
	θ_8	[1/2] 3189.13(0.03)	3381.76	0.01	232.0/1176.0
	θ_9	[2/2] 3300.93(*)	3300.93	0.00	200.0/1134.0
	θ_{10}	[7/7] 2585.16(*)	2585.16	0.00	205.1/1429.6

References

1. Abdi, A., Fukasawa, R.: On the mixing set with a knapsack constraint. *Math. Program.* **157**, 191–217 (2016)
2. Atamtürk, A.: Strong formulations of robust mixed 0–1 programming. *Math. Program.* **108**, 235–250 (2006)
3. Atamtürk, A., Nemhauser, G.L., Savelsbergh, M.W.: The mixed vertex packing problem. *Math. Program.* **89**(1), 35–53 (2000)
4. Blanchet, J., Murthy, K.: Quantifying distributional model risk via optimal transport. *Math. Oper. Res.* **44**(2), 565–600 (2019). <https://doi.org/10.1287/moor.2018.0936>
5. Calafiore, G., Campi, M.: Uncertain convex programs: randomized solutions and confidence levels. *Math. Program.* **102**, 25–46 (2005)
6. Calafiore, G., El Ghaoui, L.: On distributionally robust chance-constrained linear programs. *J. Optim. Theory Appl.* **130**, 1–22 (2006)
7. Campi, M., Garatti, S.: A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. *J. Optim. Theory Appl.* **148**, 257–280 (2011)
8. Chen, Z., Kuhn, D., Wiesemann, W.: Data-Driven Chance Constrained Programs Over Wasserstein Balls (2018). [arXiv:1809.00210](https://arxiv.org/abs/1809.00210)
9. El Ghaoui, L., Oks, M., Oustry, F.: Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Oper. Res.* **51**(4), 543–556 (2003)
10. Gao, R., Kleywegt, A.J.: Distributionally Robust Stochastic Optimization with Wasserstein Distance (2016). [arXiv:1604.02199](https://arxiv.org/abs/1604.02199)
11. Günlük, O., Pochet, Y.: Mixing mixed-integer inequalities. *Math. Program.* **90**(3), 429–457 (2001)
12. Hanasusanto, G.A., Roitch, V., Kuhn, D., Wiesemann, W.: A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Math. Program.* **151**(1), 35–62 (2015)
13. Hota, A.R., Cherukuri, A., Lygeros, J.: Data-driven chance constrained optimization under Wasserstein ambiguity sets. In: 2019 American Control Conference (ACC), pp. 1501–1506, July (2019). <https://doi.org/10.23919/ACC.2019.8814677>
14. Ji, R., and Lejeune, M.: Data-driven distributionally robust chance-constrained optimization with Wasserstein metric. Technical Report April 2019. http://www.optimization-online.org/DB_HTML/2018/07/6697.html
15. Jiang, R., Guan, Y.: Data-driven chance constrained stochastic program. *Math. Program.* **158**, 291–327 (2016)
16. Kılınç-Karzan, F., Küçükyavuz, S., Lee, D.: Joint chance-constrained programs and the intersection of mixing sets through a submodularity lens (2019). [arXiv:1910.01353](https://arxiv.org/abs/1910.01353)
17. Küçükyavuz, S.: On mixing sets arising in chance-constrained programming. *Math. Program.* **132**(1–2), 31–56 (2012)
18. Küçükyavuz, S., Noyan, N.: Cut generation for optimization problems with multivariate risk constraints. *Math. Program.* **159**(1–2), 165–199 (2016)
19. Li, B., Jiang, R., Mathieu, J.L.: Ambiguous risk constraints with moment and unimodality information. *Math. Program.* **173**(1), 151–192 (2019)
20. Liu, X., Küçükyavuz, S., Luedtke, J.: Decomposition algorithms for two-stage chance-constrained programs. *Math. Program.* **157**(1), 219–243 (2016)
21. Liu, X., Küçükyavuz, S., Noyan, N.: Robust multicriteria risk-averse stochastic programming models. *Ann. Oper. Res.* **259**(1), 259–294 (2017)
22. Liu, X., Kılınç-Karzan, F., Küçükyavuz, S.: On intersection of two mixing sets with applications to joint chance-constrained programs. *Math. Program.* **175**, 29–68 (2019)
23. Luedtke, J.: A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Math. Program.* **146**, 219–244 (2014)
24. Luedtke, J., Ahmed, S.: A sample approximation approach for optimization with probabilistic constraints. *SIAM J. Optim.* **19**, 674–699 (2008)
25. Luedtke, J., Ahmed, S., Nemhauser, G.L.: An integer programming approach for linear programs with probabilistic constraints. *Math. Program.* **122**(2), 247–272 (2010)
26. Meraklı, M., Küçükyavuz, S.: Risk aversion to parameter uncertainty in Markov decision processes with an application to slow-onset disaster relief. *IIE Trans.* **52**, 1–21 (2019). <https://doi.org/10.1080/24725854.2019.1674464>. (Article in advance)

27. Mohajerin Esfahani, P., Kuhn, D.: Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *Math. Program.* **171**(1), 115–166 (2018). <https://doi.org/10.1007/s10107-017-1172-1>
28. Noyan, N.: Risk-averse stochastic modeling and optimization. In: *TutORials in Operations Research: Recent Advances in Optimization and Modeling of Contemporary Problems*, Chapter 10, pp. 221–254. INFORMS, 2018. <https://doi.org/10.1287/educ.2018.0183>
29. Noyan, N., Meraklı, M., Küçükyavuz, S.: Two-stage stochastic programming under multivariate risk constraints with an application to humanitarian relief network design. *Math. Program.* (2019). <https://doi.org/10.1007/s10107-019-01373-4>. (**Article in advance**)
30. Rahimian, H., Mehrotra, S.: Distributionally Robust Optimization: A Review (2019). [arXiv:1908.05659](https://arxiv.org/abs/1908.05659)
31. Ruszczyński, A.: Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedra. *Math. Program.* **93**, 195–215 (2002)
32. Xie, W.: On distributionally robust chance constrained programs with Wasserstein distance. *Math. Program.* (2019). <https://doi.org/10.1007/s10107-019-01445-5>. (**Article in advance**)
33. Xie, W., Ahmed, S.: On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM J. Optim.* **28**(2), 1151–1182 (2018a)
34. Xie, W., Ahmed, S.: On quantile cuts and their closure for chance constrained optimization problems. *Math. Program.* **172**, 621–646 (2018b)
35. Zhao, M., Huang, K., Zeng, B.: A polyhedral study on chance constrained program with random right-hand side. *Math. Program.* **166**, 19–64 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Nam Ho-Nguyen¹ · Fatma Kılınc-Karzan² · Simge Küçükyavuz³ · Dabeen Lee⁴ 

✉ Dabeen Lee
dabeenl@ibs.re.kr

Nam Ho-Nguyen
nam.ho-nguyen@sydney.edu.au

Fatma Kılınc-Karzan
fkilinc@andrew.cmu.edu

Simge Küçükyavuz
simge@northwestern.edu

¹ Discipline of Business Analytics, The University of Sydney, Sydney, NSW 2006, Australia

² Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213, USA

³ Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208, USA

⁴ Discrete Mathematics Group, Institute for Basic Science (IBS), Daejeon 34126, Republic of Korea