Noname manuscript No.
(will be inserted by the editor)

Dual Decomposition of Two-Stage Distributionally Robust Mixed-Integer Programming under the Wasserstein Ambiguity Set

Kibaek Kim

Received: date / Accepted: date

Abstract We develop a dual decomposition of two-stage distributionally robust mixed-integer programming (DRMIP) under the Wasserstein ambiguity set. The dual decomposition is based on the Lagrangian dual of DRMIP, which results from the Lagrangian relaxation of the nonanticipativity constraints and min-max inequality. We present two Lagrangian dual problem formulations, each of which is based on different principle. We show that the two dual problems have the same Lagrangian bound. We develop and implement the dual decomposition method that solves the Lagrangian dual of DRMIP, which requires only minor modifications of the existing method for stochastic mixed-integer programming. We also present extensive numerical results on eighty dynamic capacity acquisition and assignment problem test instances (DCAP, one of the SIPLIB test instances) and demonstrate the computational performance of the method and the impact of the discretization properties.

Keywords dual decomposition · distributionally robust optimization · two-stage stochastic mixed-integer programming

1 Problem Statement

We consider the two-stage distributionally robust mixed-integer programming (DRMIP) problem of the form

$$z = \min_{x \in X} \left\{ c^T x + \max_{P \in \mathcal{M}} \mathbb{E}_P \left[Q(x, \omega) \right] \right\}, \tag{1a}$$

This material is based upon work supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357. The author greatly appreciates the constructive comments and suggestions from anonymous reviewers.

K. Kim
Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439
Tel.: +630-252-4832
E-mail: kimk@anl.gov

where $\mathcal{M}$ is an ambiguity set of probability distributions, the expectation is taken with respect to some probability distribution $P \in \mathcal{M}$, ω is a random evant taken from sample space Ω, and the second-stage recourse function Q is defined by

$$Q(x,\omega) := \min_{y \in Y} \left\{ b(\omega)^T y : W(\omega)y \geq h(\omega) - T(\omega)x \right\}. \tag{1b}$$

Here, $X \subseteq \mathbb{R}^{n_x}$ and $Y \subseteq \mathbb{R}^{n_y}$ can be mixed-integer sets. We assume that X is defined by a set of linear inequalities for the first-stage feasible solutions. The second-stage problem data comprise random variables $b(\omega) \in \mathbb{R}^{n_y}$, $W(\omega) \in \mathbb{R}^{m_y \times n_y}$, $h(\omega) \in \mathbb{R}^{m_y}$, and $T(\omega) \in \mathbb{R}^{m_y \times n_x}$ for a given realization $\omega \in \Omega$. DRMIP is a distributionally robust variant of two-stage stochastic mixed-integer programming (SMIP), which accounts for the worst-case realization of the underlying distribution for the second-stage uncertainty. Therefore, DRMIP offers a statistically robust solution to the uncertainty of the empirical distribution.

In this paper, we focus on the Wasserstein ambiguity set, which is defined as a set of distributions around the empirical distribution. While reformulation approaches have been actively studied (e.g., [11,32,31]), only a few studies have developed numerical algorithms for solving two-stage distributionally robust optimization under the Wasserstein (or generic) ambiguity set (e.g., [5,24,6,30]), of which even fewer studies consider discrete variables in the first and/or second stage. For example, a Benders decomposition method has been developed with a distribution separation procedure for solving two-stage DRMIP with the Wasserstein ambiguity set [5]. A disjunctive programming method, developed for SMIP [29], has been adapted to solve DRMIP within the Benders decomposition framework [24]. However, the class of two-stage DRMIP addressed in [5,24] is limited to having pure-binary variables in the first stage. As in SMIP, the presence of first-stage continuous variables and second-stage integer variables makes DRMIP extremely challenging, because the recourse function is nonconvex and discontinuous in the first-stage variables and also because the first-stage feasible solutions are infinitely many.

The dual decomposition method was developed by Carøe and Schultz [8] and has been used successfully for solving SMIP (e.g., [19,16,20,1,26,4,14,23]) as well as other structured optimization problems (e.g., [27,33,9,10]). The dual decomposition is obtained by applying a Lagrangian relaxation of the nonanticipativity constraints on the first-stage variables. Most important, the resulting Lagrangian dual problem is decomposed for each scenario and can be solved in parallel. Since the Lagrangian dual problem is a single-scenario deterministic equivalent problem, the dual decomposition can naturally address mixed-integer variables in the first and second stages by using state-of-the-art MIP algorithms and solvers. Moreover, an open-source software package DSP [20] enables researchers to prototype, test, and develop new algorithmic features for SMIP. For example, the dual decomposition has been advanced to asynchronous parallel computation [18] and a global optimization solver with branch-and-bound method [17].

In this paper, we develop a dual decomposition method for solving DRMIP under the Wasserstein ambiguity set $\mathcal{M}$. To this end, we derive a deterministic equivalent problem of DRMIP as a large-scale deterministic MIP problem for a given finite support. We present two different formulations of the Lagrangian dual of DRMIP, which leads to the dual decomposition. One formulation is obtained by applying the Lagrangian relaxation of the nonanticipativity constraints on the first-stage variables and the min-max inequality, while the other is obtained by considering the deterministic equivalent problem as a SMIP. We then present the dual decomposition method that can be implemented with minor modifications to that for SMIP. We also perform extensive numerical experiments to demonstrate the computational performance of the dual decomposition and the impact of the discretization properties.

The remainder of this paper is organized as follows. In Section 3 we present the deterministic equivalent formulation of DRMIP. Section 4 presents the Lagrangian dual problems. The dual decomposition method is developed in Section 5. In Section 6 we present the implementation of the method in DSP and numerical experiments. In Section 7 we summarize the paper and discuss directions of future work.

2 Preliminaries and Assumptions

We assume that each probability distribution $P \in \mathcal{M}$ has finite support Ω such that $|\Omega|$ is finite. For each $\omega \in \Omega$, we define a random vector $\xi(\omega)^T := [b(\omega)^T, h(\omega)^T, W_1(\omega), \ldots, W_{m_y}(\omega), T_1(\omega), \ldots, T_{m_y}(\omega)] \in \mathbb{R}^{n_\omega}$. The Wasserstein distance of order p between P and $\hat{P}$ is defined by

Do we need to define the dimension?

$$W_p(P, \hat{P}) := \left(\min_{\pi \in \Pi(P,\hat{P})} \sum_{\omega \in \Omega} \sum_{\hat{\omega} \in \Omega} \|\xi(\omega) - \xi(\hat{\omega})\|^p \pi(\xi(\omega), \xi(\hat{\omega})) \right)^{1/p}, \qquad (2)$$

where $\Pi(P, \hat{P})$ is the set of all joint probability distributions of $\xi(\omega)$ and $\xi(\hat{\omega})$ with marginals P and $\hat{P}$, and $\|\cdot\|$ is a norm on $\mathbb{R}^{n_\omega}$. Then, for $\epsilon > 0$, the Wasserstein ambiguity set is defined as the Wasserstein ball of probability distributions

$$\mathcal{P} := \left\{ P \in \mathcal{M}(\Omega) : W_p(P, \hat{P}) \leq \epsilon \right\}. \qquad (3)$$

Let $\{\hat{\omega}_1, \ldots, \hat{\omega}_N\}$ be the set of empirical observations with the corresponding probability estimates $\hat{p}_1, \ldots, \hat{p}_N$. We denote $\hat{P}_N = \sum_{s=1}^{N} \hat{p}_s \delta_{\hat{\omega}_s}$. We define the ambiguity set of distributions based on the Wasserstein metric at the em-

pirical observations as

$$\mathcal{P} := \left\{ P \in \mathcal{M} : \begin{array}{l} \sum_{\omega\in\Omega}\sum_{s=1}^{N} u_s(\omega)\|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \le \epsilon^p, \\ \sum_{\omega\in\Omega} u_s(\omega) = \hat{p}_s \quad \forall s = 1,\dots,N, \\ \sum_{s=1}^{N} u_s(\omega) = P(\omega) \quad \forall \omega \in \Omega, \\ u_s(\omega) \ge 0 \quad \forall \omega \in \Omega,\ s = 1,\dots,N \end{array} \right\}, \tag{4}$$

where ϵ is the Wasserstein distance limit.

3 Deterministic Equivalent Formulation

We present the deterministic equivalent formulation of DRMIP, which is of large-scale MIP. As a special case of Theorem 1 in [13] for finite support, the following duality property is key to deriving our reformulation.

Lemma 1 (Theorem 1 in [13]) *Suppose that $\mathcal{P}$ is as defined in (4). Consider any function f defined on Ω such that $f(\hat{\omega}_s) < \infty$ for $s = 1,\dots,N$. Then, the strong duality property holds for the following linear programming problem:*

$$\max_{P\in\mathcal{P}} \mathbb{E}_P\left[f(\omega)\right]. \tag{5}$$

Furthermore, its dual is given as the following linear program:

$$\min_{\alpha\ge 0,\beta_s} \quad \epsilon^p\alpha + \sum_{s=1}^{N} \hat{p}_s\beta_s \tag{6a}$$

$$s.t. \quad \|\xi(\hat{\omega}_s) - \xi(\omega)\|^p\alpha + \beta_s \ge f(\omega) \quad \forall \omega \in \Omega,\ s = 1,\dots,N. \tag{6b}$$

Proof We observe that the maximization problem is a linear programming problem. Consider problem (5) given by

$$\max_{P(\omega),u_s(\omega)} \quad \sum_{\omega\in\Omega} P(\omega)f(\omega) \tag{7a}$$

$$\text{s.t.} \quad \sum_{\omega\in\Omega}\sum_{s=1}^{N} u_s(\omega)\|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \le \epsilon^p, \quad (\alpha) \tag{7b}$$

$$\sum_{\omega\in\Omega} u_s(\omega) = \hat{p}_s \quad \forall s = 1,\dots,N, \quad (\beta_s) \tag{7c}$$

$$\sum_{s=1}^{N} u_s(\omega) = P(\omega) \quad \forall \omega \in \Omega, \tag{7d}$$

$$\sum_{\omega\in\Omega} P(\omega) = 1, \tag{7e}$$
$$P(\omega) \geq 0 \quad \forall \omega \in \Omega, \tag{7f}$$
$$u_s(\omega) \geq 0 \quad \forall s = 1,\dots,N,\ \omega \in \Omega. \tag{7g}$$

By substituting $P(\omega)$ with $\sum_{s=1}^{N} u_s(\omega)$ from (7d), the problem can be reformulated as the following form,

$$\max_{u_s(\omega)} \quad \sum_{\omega\in\Omega}\sum_{s=1}^{N} u_s(\omega) f(\omega) \tag{8a}$$
$$\text{s.t.} \quad (7b) - (7c), (7g),$$
$$\sum_{\omega\in\Omega}\sum_{s=1}^{N} u_s(\omega) = 1, \tag{8b}$$
$$\sum_{s=1}^{N} u_s(\omega) \geq 0 \quad \forall \omega \in \Omega. \tag{8c}$$

Note that constraints (8b) and (8c) are implied by constraints (7c) and (7g), respectively, and can further be removed from the problem (8). Let (α, β_s) be the Lagrangian multipliers with respect to the constraints. Then, the dual of (8) is given in (6). □

The strong duality result allows us to derive the reformulation of two-stage DRMIP as follows.

Proposition 1 *Suppose that $\mathcal{P}$ is defined by (4). Two-stage DRMIP (1) can be written as a deterministic MIP of the form*

$$\min_{x,\alpha,\beta_s,y(\omega)} \quad c^T x + \epsilon^p \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s \tag{9a}$$
$$s.t. \quad \|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \alpha + \beta_s - d(\omega)^T y(\omega) \geq 0 \quad \forall \omega \in \Omega,\ s = 1,\dots,N, \tag{9b}$$
$$T(\omega)x + W(\omega)y(\omega) \geq h(\omega) \quad \forall \omega \in \Omega, \tag{9c}$$
$$x \in X, \quad \alpha \geq 0, \quad y(\omega) \in Y \quad \forall \omega \in \Omega. \tag{9d}$$

Proof The problem (1) can be written as the following problem:

$$\min_{x\in X, r\in\mathbb{R}, y(\omega)\in Y} \quad c^T x + r \tag{10a}$$
$$\text{s.t.} \quad r \geq \mathbb{E}_P\left[d(\omega)^T y(\omega)\right] \quad \forall P \in \mathcal{P}, \tag{10b}$$
$$T(\omega)x + W(\omega)y(\omega) \geq h(\omega) \quad \forall \omega \in \Omega. \tag{10c}$$

The constraint (10b) is equivalent to $r \geq \max_{P\in\mathcal{P}} \mathbb{E}_P\left[d(\omega)^T y(\omega)\right] =$ (6) by Lemma 1, and thus

$$r \geq \epsilon^p \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s \tag{11a}$$

$$\|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \alpha + \beta_s \geq d(\omega)^T y(\omega) \quad \forall \omega \in \Omega,\ s = 1, \ldots, N. \tag{11b}$$

Substituting variable r in (10b) and (11a) results in (9). □

The MIP reformulation (9) embeds the block-angular structure that is observed in two-stage SMIP (or two-stage robust optimization), in which variables (x, α, β_s) can be considered as the first-stage (i.e., nonanticipative) variables. Specifically, problem (9) can be written as

$$\min_{(x,\alpha,\beta)\in\mathcal{X}} \left\{ g(x,\alpha,\beta) := c^T x + \epsilon^p \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s + \sum_{\omega\in\Omega} G(x,\alpha,\beta,\omega) \right\},$$

where $\mathcal{X} := X \times \mathbb{R}_+ \times \mathbb{R}^N$ is the feasible set of first-stage variables, $\beta := (\beta_1, \ldots, \beta_N)$, and

$$\begin{aligned} G(x,\alpha,\beta,\omega) := \min_{y(\omega)} \quad & 0 \\ \text{s.t.} \quad & d(\omega)^T y(\omega) \leq \|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \alpha + \beta_s \quad \forall s = 1, \ldots, N, \\ & W(\omega) y(\omega) \geq h(\omega) - T(\omega) x, \\ & y(\omega) \in Y. \end{aligned}$$

4 Lagrangian Dual of DRMIP

It may be impractical to solve the deterministic equivalent problem particularly with a large finite support set Ω. In this section, we present the Lagrangian dual of DRMIP, which would enable us to use the existing dual decomposition methods (e.g., [18,23,17,26]) with only minor modifications. Similar to that of SMIP, the Lagrangian dual is based on the Lagrangian relaxation with respect to the nonanticipativity constraints.

Specifically, we present two approaches for deriving the Lagrangian dual of DRMIP. One approach is developed by applying the mix-max inequality. This approach is independent of the definition of the ambiguity set and can be seen as a direct extension from the dual decomposition of SMIP. The other approach is developed for the Wasserstein-based ambiguity set, as defined in (4), by applying the Lagrangian relaxation technique used for SMIP. In particular, the problem (9) can be seen as a SMIP, where (x, α, β_s) are the first-stage variables and $y(\omega)$ is the second-stage variable. Most interesting, we show that the Lagrangian dual problems resulting from the two different approaches are equivalent. We first derive the Lagrangian dual of the problem (1) by using the min-max inequality.

Proposition 2 *The Lagrangian dual of problem (1) is given by*

$$z_{LD} = \max_{\lambda(\omega), P\in\mathcal{P}} \quad \sum_{\omega\in\Omega} P(\omega) D(\lambda(\omega), \omega) \tag{12a}$$

$$s.t. \quad \sum_{\omega\in\Omega} P(\omega)\lambda(\omega) = c, \tag{12b}$$

where

$$D(\lambda(\omega),\omega) = \min_{x\in X, y\in Y} \quad \lambda(\omega)^T x + d(\omega)^T y \tag{13a}$$

$$s.t. \quad T(\omega)x + W(\omega)y \geq h(\omega). \tag{13b}$$

Proof The proof is given by applying the min-max inequality. We consider the equivalent formulation of (1) as

$$\min_{x,x(\omega)} \max_{P\in\mathcal{P}} \quad \mathbb{E}_P\left[c^T x + Q(x(\omega),\omega)\right] \tag{14a}$$

$$\text{s.t.} \quad [x - x(\omega)]\, P(\omega) = 0 \quad \forall\omega\in\Omega, \tag{14b}$$

$$x(\omega) \in X \quad \forall\omega\in\Omega, \tag{14c}$$

where (14b) are called the nonanticivativity constraints. By taking the Lagrangian relaxation with respect to (14b), we have the following Lagrangian function:

$$\min_{x,x(\omega)\in X} \max_{P\in\mathcal{P}} \mathbb{E}_P\left[(c - \lambda(\omega))^T x + \lambda(\omega)^T x(\omega) + Q(x(\omega),\omega)\right],$$

where $\lambda(\omega)$ are the Lagrangian multipliers corresponding to the constraints (14b). For the Lagrangian function value to be bounded, we need

$$\mathbb{E}_P[\lambda(\omega)] = c, \tag{15}$$

and thus the corresponding term in the objective function is eliminated. Hence, the objective function is written as

$$\max_{P\in\mathcal{P}} \mathbb{E}_P\left[\lambda(\omega)x(\omega) + Q(x(\omega),\omega)\right].$$

In addition, by the min-max inequality, we have

$$\min_{x(\omega)\in X} \max_{P\in\mathcal{P}} \mathbb{E}_P\left[\lambda(\omega)x(\omega) + Q(x(\omega),\omega)\right] \geq \max_{P\in\mathcal{P}} \min_{x(\omega)\in X} \mathbb{E}_P\left[\lambda(\omega)x(\omega) + Q(x(\omega),\omega)\right].$$

With constraint (15), we have the dual problem (12). □

Note that the weak duality holds; that is, $z(\Omega) \geq z_{LD}(\Omega)$. The Lagrangian dual problem (12) can be seen as a direct extension from SMIP to DRMIP and is equivalent to that of SMIP for a singleton $\mathcal{P}$ (i.e., when the distribution is known). As in the dual decomposition of SMIP, evaluating the Lagrangian dual function D can be parallelized for each $\omega \in \Omega$. In addition, regardless of the specification of $\mathcal{P}$, the scenario subproblems $D(\lambda(\omega),\omega)$ are equivalent to those for SMIP. The Lagrangian dual (12) of DRMIP differs from that of SMIP only by perturbing the probability measure defined in $\mathcal{P}$.

However, the Lagrangian dual problem (12) can be nonlinear even with the piecewise linear concave function D because of the bilinear term in the objective function. The bilinear term can linearized as follows.

Corollary 1 *The Lagrangian dual problem* (12) *is equivalent to*

$$z_{LD} = \max_{\mu(\omega), P \in \mathcal{P}} \quad \sum_{\omega \in \Omega} \bar{D}(\mu(\omega), P(\omega), \omega) \tag{16a}$$

$$s.t. \quad \sum_{\omega \in \Omega} \mu(\omega) = c, \tag{16b}$$

where

$$\bar{D}(\mu(\omega), P(\omega), \omega) := \min_{x \in X, y \in Y} \quad \mu(\omega)^T x + P(\omega) d(\omega)^T y \tag{17a}$$

$$s.t. \quad T(\omega)x + W(\omega)y \geq h(\omega). \tag{17b}$$

Proof The equivalent formulation (16) is obtained by substituting $\lambda(\omega)P(\omega)$ by $\mu(\omega)$ in (12) and (13). □

We now present the Lagrangian dual of the deterministic equivalent DRMIP (9) with the Wasserstein ambiguity set. To this end, we consider the following equivalent formulation of DRMIP (9):

$$\min_{x, \alpha, \beta_s, x(\omega), \alpha(\omega), \beta_s(\omega)} \quad c^T x + \epsilon^p \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s \tag{18a}$$

$$\text{s.t. } x - x(\omega) = 0 \quad \forall \omega \in \Omega, \tag{18b}$$

$$\alpha - \alpha(\omega) = 0 \quad \forall \omega \in \Omega, \tag{18c}$$

$$\beta_s - \beta_s(\omega) = 0 \quad \forall \omega \in \Omega,\ s = 1, \ldots, N, \tag{18d}$$

$$\|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \alpha(\omega) + \beta_s(\omega) \geq Q(x, \omega) \quad \forall \omega \in \Omega,\ s = 1, \ldots, N, \tag{18e}$$

$$\alpha(\omega) \geq 0 \quad \forall \omega \in \Omega, \tag{18f}$$

where (18b)–(18d) are the nonanticipativity constraints. Note that the reformulation technique used in (18) has been commonly used to derive the dual or scenario decomposition of SMIP (e.g., [8,1,20]).

The following equivalent formulation relaxes the nonanticipativity constraints (18c) and (18d) to inequalities, which may benefit to the dual decomposition method by reducing the size of dual search space.

Proposition 3 *DRMIP* (9) *is equivalent to*

$$\min_{x, \alpha, \beta_s, x(\omega), \alpha(\omega), \beta_s(\omega), y(\omega)} \quad c^T x + \epsilon^p \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s \tag{19a}$$

$$s.t. \quad x - x(\omega) = 0 \quad \forall \omega \in \Omega, \tag{19b}$$

$$\alpha - \alpha(\omega) \geq 0 \quad \forall \omega \in \Omega, \tag{19c}$$

$$\beta_s - \beta_s(\omega) \geq 0 \quad \forall \omega \in \Omega,\ s = 1, \ldots, N, \tag{19d}$$

$$\|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \alpha(\omega) + \beta_s(\omega) - d(\omega)^T y(\omega) \geq 0 \quad \forall \omega \in \Omega,\ s = 1, \ldots, N, \tag{19e}$$

$$T(\omega)x(\omega) + W(\omega)y(\omega) \geq h(\omega) \quad \forall \omega \in \Omega, \tag{19f}$$
$$x(\omega) \in X, \quad \alpha(\omega) \geq 0, \quad y(\omega) \in Y \quad \forall \omega \in \Omega$$
$$\forall \omega \in \Omega,\ s = 1, \ldots, N. \tag{19g}$$

Proof We rewrite the problem (18) by introducing slack variables to constraints (18e) as follows:

$$\min_{\substack{x,\alpha,\beta_s,\\ x(\omega),\alpha(\omega),\beta_s(\omega),\\ \alpha'(\omega),\beta'_s(\omega),\gamma_s(\omega)}} \quad c^T x + \epsilon^p \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s \tag{20a}$$
$$\text{s.t. } (18b) - (18d), (18f)$$
$$\|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \left(\alpha(\omega) - \alpha'(\omega)\right) + \left(\beta_s(\omega) - \beta'_s(\omega)\right) - \gamma_s(\omega)$$
$$= Q(x, \omega) \quad \forall \omega \in \Omega,\ s = 1, \ldots, N, \tag{20b}$$
$$\alpha(\omega) \geq \alpha'(\omega) \geq 0,\ \beta'_s(\omega) \geq 0,\ \gamma_s(\omega) \geq 0, \tag{20c}$$

where $\alpha'(\omega), \beta'(\omega), \gamma_s(\omega)$ are the nonnegative slack variables. By substituting $\bar{\alpha}(\omega) := \alpha(\omega) - \alpha'(\omega)$ and $\bar{\beta}_s(\omega) := \beta_s(\omega) - \beta'_s(\omega)$, the constraints of problem (20) can be written as

$$x - x(\omega) = 0 \quad \forall \omega \in \Omega,$$
$$\alpha - \bar{\alpha}(\omega) - \alpha'(\omega) = 0 \quad \forall \omega \in \Omega,$$
$$\beta_s - \bar{\beta}_s(\omega) - \beta'_s(\omega) = 0 \quad \forall \omega \in \Omega,\ s = 1, \ldots, N,$$
$$\|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \bar{\alpha}(\omega) + \bar{\beta}_s(\omega) - \gamma_s(\omega) = Q(x, \omega) \quad \forall \omega \in \Omega,\ s = 1, \ldots, N,$$
$$\bar{\alpha}(\omega) \geq 0,\ \alpha'(\omega) \geq 0,\ \beta'_s(\omega) \geq 0,\ \gamma_s(\omega) \geq 0 \quad \forall \omega \in \Omega,\ s = 1, \ldots, N,$$

or equivalently,

$$x - x(\omega) = 0 \quad \forall \omega \in \Omega,$$
$$\alpha - \bar{\alpha}(\omega) \geq 0 \quad \forall \omega \in \Omega,$$
$$\beta_s - \bar{\beta}_s(\omega) \geq 0 \quad \forall \omega \in \Omega,\ s = 1, \ldots, N,$$
$$\|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \bar{\alpha}(\omega) + \bar{\beta}_s(\omega) \geq Q(x, \omega) \quad \forall \omega \in \Omega,\ s = 1, \ldots, N,$$
$$\bar{\alpha}(\omega) \geq 0 \quad \forall \omega \in \Omega.$$

This results in the formulation as in (19). □

For the rest of the paper, we use the deterministic equivalent formulation as in (19).

Proposition 4 *The Lagrangian relaxation of* (19) *with respect to the nonanticipativity constraints* (19b), (19c), *and* (19d) *gives the following Lagrangian dual problem:*

$$z_{WLD} = \max_{\mu(\omega), \nu(\omega) \geq 0, u_s(\omega) \geq 0} \quad \sum_{\omega \in \Omega} D_W(\mu(\omega), \nu(\omega), u(\omega), \omega) \tag{21a}$$

$$s.t. \quad \sum_{\omega\in\Omega} \mu(\omega) = c, \tag{21b}$$
$$\sum_{\omega\in\Omega} \nu(\omega) = \epsilon^p, \tag{21c}$$
$$\sum_{\omega\in\Omega} u_s(\omega) = \hat{p}_s \quad \forall s = 1,\dots,N, \tag{21d}$$

where $u(\omega) := (u_1(\omega),\dots,u_N(\omega))$ *is defined for* $\omega \in \Omega$ *and the Lagrangian function* $D_W(\mu(\omega),\nu(\omega),u(\omega),\omega)$ *is defined for each* $\omega \in \Omega$ *as*

$$\min_{x,\alpha,\beta_s,y} \quad \mu(\omega)^T x + \nu(\omega)\alpha + \sum_{s=1}^{N} u_s(\omega)\beta_s \tag{22a}$$
$$s.t. \quad \|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \alpha + \beta_s - d(\omega)^T y \geq 0 \quad \forall s = 1,\dots,N, \tag{22b}$$
$$T(\omega)x + W(\omega)y \geq h(\omega), \tag{22c}$$
$$x \in X, \quad \alpha \geq 0, \quad y \in Y. \tag{22d}$$

Proof Let $\mu(\omega)$, $\nu(\omega) \geq 0$, and $u_s(\omega) \geq 0$ be the Lagrangian multipliers associated with the nonanticipativity constraints (19b), (19c), and (19d), respectively. The Lagrangian relaxation of the nonanticipativity constraints gives the Lagrangian dual function of the form

$$\begin{aligned}\min_{x,\alpha,\beta_s,x(\omega),\alpha(\omega),\beta_s(\omega),y(\omega)} \quad & \left(c - \sum_{\omega\in\Omega}\mu(\omega)\right)^T x + \sum_{\omega\in\Omega}\mu(\omega)x(\omega) \\ & + \left(\epsilon^p - \sum_{\omega\in\Omega}\nu(\omega)\right)\alpha + \sum_{\omega\in\Omega}\nu(\omega)\alpha(\omega) \\ & + \sum_{s=1}^{N}\left(\hat{p}_s - \sum_{\omega\in\Omega} u_s(\omega)\right)\beta_s + \sum_{\omega\in\Omega}\sum_{s=1}^{N} u_s(\omega)\beta_s(\omega) \\ \text{s.t.} \quad & (19e) - (19g).\end{aligned}$$

In order to ensure the boundedness of the Lagrangian dual function value, the Lagrangian multipliers are subject to

$$\sum_{\omega\in\Omega}\mu(\omega) = c, \quad \sum_{\omega\in\Omega}\nu(\omega) = \epsilon^p, \quad \sum_{\omega\in\Omega} u_s(\omega) = \hat{p}_s \quad \forall s = 1,\dots,N. \tag{23}$$

As a result, the variables x, α, and β_s are removed from the problem, and thus the Lagrangian dual function is given by

$$\begin{aligned}\min_{x(\omega),\alpha(\omega),\beta_s(\omega),y(\omega)} \quad & \sum_{\omega\in\Omega}\left[\mu(\omega)x(\omega) + \nu(\omega)\alpha(\omega) + \sum_{s=1}^{N} u_s(\omega)\beta_s(\omega)\right] \\ \text{s.t.} \quad & (19e) - (19g),\end{aligned}$$

which can be decomposed for each $\omega \in \Omega$. With the boundedness conditions (23), we have the Lagrangian dual problem as in (21). □

We highlight that Proposition 4 is independent to the choice of Wasserstein norm in (22b). This will provide the additional flexibility to the modeling of DRMIP in practice. We note that the Lagrangian dual problem (21) may be computationally more expensive than the problem (12). Specifically, each scenario subproblem (22) of the new dual problem (21) has $N+1$ variables and N constraints more than those of (12). Nevertheless, the Lagrangian dual problem (21) results in the same Lagrangian bound as the problem (12) does.

Theorem 1 *Suppose that $\mathcal{P}$ is defined as (4). $z_{WLD} = z_{LD}$.*

Proof Introducing variables r_s, the scenario subproblem (22) for a given $\omega \in \Omega$ is equivalent to

$$\begin{aligned} \min_{x,\alpha,\beta_s,r_s,y} \quad & \mu(\omega)^T x + \nu(\omega)\alpha + \sum_{s=1}^{N} u_s(\omega)\beta_s \\ \text{s.t.} \quad & \|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \alpha + \beta_s - d(\omega)^T y - r_s = 0 \quad \forall s = 1,\dots,N, \\ & (22c) - (22d),\ r_s \geq 0 \quad \forall s = 1,\dots,N. \end{aligned}$$

Eliminating variables β_s from this equivalent problem, we have

$$\begin{aligned} \min_{x,\alpha,\beta_s,r_s,y} \quad & \mu(\omega)^T x + \left[\nu(\omega) - \sum_{s=1}^{N} u_s(\omega)\|\xi(\hat{\omega}_s) - \xi(\omega)\|^p\right]\alpha \\ & + \sum_{s=1}^{N} u_s(\omega) d(\omega)^T y + \sum_{s=1}^{N} u_s(\omega) r_s \\ \text{s.t.} \quad & (22c) - (22d),\ r_s \geq 0 \quad \forall s = 1,\dots,N. \end{aligned}$$

We observe that $\alpha = 0$ if

$$\nu(\omega) - \sum_{s=1}^{N} u_s(\omega)\|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \geq 0. \tag{24}$$

Otherwise, the subproblem is unbounded. In addition, $r_s = 0$ since $u_s(\omega) \geq 0$ for all $s = 1,\dots,N$. Therefore, the scenario subproblem (22) is equivalent to

$$\min_{x,y} \left\{ \mu(\omega)^T x + \sum_{s=1}^{N} u_s(\omega) d(\omega)^T y : (22c) - (22d), (24) \right\},$$

and thus also equivalent to the scenario subproblem (17) by substituting $\sum_{s=1}^{N} u_s(\omega)$ with $P(\omega)$; that is,

$$P(\omega) = \sum_{s=1}^{N} u_s(\omega) \quad \forall \omega \in \Omega. \tag{25}$$

By adding the constraints (24) and (25) to the problem (21), we have the Lagrangian dual problem (16). □

A key advantage of using the Lagrangian dual problem (21) may be that existing scenario decomposition methods and solvers for SMIP can be used without any modification (e.g. [1,26,17]). Note, however, that scenario subproblems can be unbounded for some dual variable values, particularly if inequality (24) is violated. The unbounded solutions may be avoided by introducing arbitrarily large bounds to α and β_s. In our numerical experiments, however, we found that the dual decomposition method implemented in DSP suffered from numerical instability unless good (tight) bounds are set for α and β_s. On the other hand, only minor modification to existing SMIP solvers would be required for solving the Lagrangian dual problem (16).

5 Dual Decomposition

We present the dual decomposition method that can be applied to the Lagrangian dual problem (16). We adapt the dual decomposition method that has been developed for SMIP in [8,20] with the additional dual constraints defined by $\mathcal{P}$.

A main idea of the dual decomposition is the outer approximation of the Lagrangian dual function $\bar{D}(\mu(\omega), P(\omega), \omega)$ by a set of linear inequalities (i.e., subgradients), which has been developed to bundle methods. The bundle methods have been particularly successful in the dual decomposition of SMIP, because the methods guarantee the finite termination at optimum, as compared with subgradient methods that require tuning step size. Extensive numerical experiments are also available in [20]. The bundle methods are also capable of easily incorporating additional constraints from the ambiguity set $\mathcal{P}$ as well as regularization (e.g., [20,18]).

After t iterations, the outer approximation of the Lagrangian dual problem (16) is given by

$$
\begin{aligned}
z_t := \max_{\theta(\omega),\mu(\omega),P(\omega)} \quad & \sum_{\omega\in\Omega} \theta(\omega) && \text{(26a)}\\
\text{s.t.} \quad & \sum_{\omega\in\Omega} \mu(\omega) = c, && \text{(26b)}\\
& (P(\omega))_{\omega\in\Omega} \in \mathcal{P}, && \text{(26c)}\\
& \theta(\omega) \le \hat{x}_j(\omega)^T \mu(\omega) + d(\omega)^T \hat{y}_j(\omega) P(\omega) \\
& \qquad \forall j = 1,\dots,t,\ \omega \in \Omega, && \text{(26d)}\\
& \|\mu(\omega) - \hat{\mu}_t(\omega)\| \le \Delta_t \quad \forall \omega \in \Omega, && \text{(26e)}\\
& \|P(\omega) - \hat{P}_t(\omega)\| \le \Delta_t \quad \forall \omega \in \Omega, && \text{(26f)}
\end{aligned}
$$

where $\mathcal{P}$ is defined as

$$\mathcal{P} := \left\{ P \in \mathcal{M} : \begin{array}{l} \sum_{\omega\in\Omega}\sum_{s=1}^{N} u_s(\omega)\|\xi(\hat{\omega}_s) - \xi(\omega)\|^p \le \epsilon^p, \\ \sum_{\omega\in\Omega} u_s(\omega) = \hat{p}_s \quad \forall s = 1,\dots,N, \\ \sum_{s=1}^{N} u_s(\omega) = P(\omega) \quad \forall \omega \in \Omega, \\ u_s(\omega) \ge 0 \quad \forall \omega \in \Omega,\ s = 1,\dots,N \end{array} \right\}, \tag{27}$$

the right-hand side of constraints (26d) are the subgradients of $\bar{D}$ at each iteration $j = 1,\dots,t$ for each $\omega \in \Omega$; and constraints (26e) and (26f) are the trust-region (TR) constraints with center $(\hat{\mu}_t(\omega), \hat{P}_t(\omega))$ and size Δ_t at iteration t.

The problem (26) is a large-scale linear program with $|\Omega|(n_x + N + 2)$ variables and $|\Omega| + n_x + N + 1$ constraints, excluding the outer-approximation cuts (26d) and TR column bounds (26e) and (26f). We remark that the sequence of linear programs can be efficiently resolved with warm-starting by off-the-shelf solvers. Note, however, a decomposition approach can also be employed for extremely large problems. An example of the decomposition can be obtained by iteratively generating $P(\omega)$ with respect to the linear cuts (26d), where the corresponding cut generation solves a linear program over $\mathcal{P}$.

We dynamically adjust the TR center and size in order to regularize the iterates. Let $\underline{\Delta}$ and $\bar{\Delta}$ be the minimum and maximum values of the TR size Δ_t, respectively. We denote by $(\hat{x}_t(\omega), \hat{y}_t(\omega))$ an optimal solution of the Lagrangian subproblem (17) for $(\mu_t(\omega), P_t(\omega))$ at iteration t. We also denote by $\bar{D}_t := \sum_{\omega\in\Omega} \bar{D}(\mu_t(\omega), P_t(\omega), \omega)$ the Lagrangian dual value at iteration t. We update the TR center if the Lagrangian function value is sufficiently increased at iteration t as

$$\bar{D}_t \ge \hat{D}_t + \eta\zeta_t, \tag{28}$$

where $\eta \in (0, 0.5)$, $\hat{D}_t$ is the Lagrangian dual value at the TR center, and $\zeta_t := z_t - \hat{D}_t$ (i.e., predicted increase). Then, the bundle master problem (26) is solved with the updated TR. This iteration is called a *serious step*. Otherwise, we call the iteration a *null step*, where a new set of cuts (26d) may be added to improve the outer approximation.

A careful update of the TR size Δ_t is significant for the performance (i.e., number of iterations). For example, if the size is too large, the bundle master may suffer from oscillating the solutions and take a number of unnecessary null steps before each serious step is taken. On the other hand, if the size is too small, the bundle master may take serious steps with marginal decreases of the Lagrangian function value. We adapt the update procedure developed in [18] for the TR size Δ_t. We define the approximation error measured by

$$\rho := \min\left\{1, \left\| \begin{array}{l} \mu_t(\omega) - \hat{\mu}_t(\omega) \\ P_t(\omega) - \hat{P}_t(\omega) \end{array} \right\|\right\} \frac{\max\{\hat{D}_t - \bar{D}_t, l_t\}}{\zeta_t}, \tag{29}$$

and the model linearization error measured by

$$l_t := \sum_{\omega \in \Omega_t} \left[\hat{x}_t(\omega)^T \hat{\mu}_t(\omega) + d(\omega)^T \hat{y}_t(\omega) \hat{P}_t(\omega) \right] - \hat{D}_t. \tag{30}$$

The TR size is updated as follows:

1. If $\rho > 0$, then set $\tau_{t+1} \leftarrow \tau_t + 1$.
2. If $\rho > \underline{\rho}$ or $(\rho \in (0, \underline{\rho})$ and $\tau_t \geq \bar{\rho})$, then set $\tau_{t+1} \leftarrow 0$ and

$$\Delta_{t+1} \leftarrow \max \left\{ \frac{\Delta_t}{\min\{\rho, \underline{\rho}\}}, \bar{\Delta} \right\}. \tag{31}$$

Here τ_t counts the number of iterations in which TR is not reduced, and $\underline{\rho}$ and $\bar{\rho}$ are given parameters. On the other hand, we may detect that the TR is too small if the solution is bounded by the TR and if $\hat{D}_{t+1} \geq \hat{D}_t + 0.5\zeta_t$. Then, we increase the TR by $\Delta_{t+1} \leftarrow \min\{2\Delta_t, \bar{\Delta}\}$. More details of the algorithm development (e.g., serious and null steps, TR updates) are available in [22,18].

We have implemented an algorithm to find upper bounds for first-stage solution $x_t(\omega)$. Suppose that the first-stage solutions $x_t(\omega)$ are given for $\omega \in \Omega$ at iteration t. We evaluate the recourse function $Q(x_t(\omega), \omega)$ for each $\omega \in \Omega$. Note that each iteration can have at most k different first-stage solutions for the evaluation and that each evaluation can be parallelized as described in [20]. Then, an upper bound can be computed at the first-stage solution x_t by solving the following linear programming problem:

$$\bar{z}(x_t) := c^T x_t + \min_{\alpha \geq 0, \beta_s} \quad \epsilon\alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s \tag{32a}$$

$$\text{s.t.} \quad \|\hat{\omega}_s - \omega\|\alpha + \beta_s \geq Q(x_t, \omega) \quad \forall \omega \in \Omega,\ s = 1, \ldots, N. \tag{32b}$$

We summarize the algorithmic steps of the dual decomposition method for DRMIP as follows.

Algorithm 1 is adapted from the BTR method [18] that implements a bundle TR method for SMIP. The algorithm initializes a dual feasible solution $(\mu_0(\omega), P_0(\omega))$ for all $\omega \in \Omega$ and other parameters, including optimality gap δ and serious-step condition η (see lines 12 and 15, respectively). The initial steps of the algorithm include the Lagrangian subproblem solution (line 2) and generating outer-approximation cuts (line 6). We emphasize that the Lagrangian subproblem solutions (lines 2 and 10) can be parallelized by utilizing multiple CPUs. We also compute the upper bounds for some $x_t(\omega)$ obtained from the subproblems (lines 3 and 11). The iterative steps (lines 8–24) are taken by solving the bundle master (line 9) and the subproblems (line 10) until the optimality condition (line 12) is satisfied. In addition, the serious steps are taken in lines 16–18 in order to update the best dual bound $\hat{D}_{t+1}$ (line 16) and the TR (lines 17 and 18). The TR size may be increased to aggressively search for dual variable values. When the null steps are taken in lines 20–23, the outer approximation cuts are added, and the TR size may be decreased. We follow the TR updates as described above.

Algorithm 1 A dual decomposition method for DRMIP

Initialize a feasible $(\mu_0(\omega), P_0(\omega))$ for $\omega \in \Omega$, $\Delta_0 \in [\underline{\Delta}, \bar{\Delta}]$, $\delta \geq 0$, $\eta \in (0, 0.5)$, $z_{UB} \leftarrow \infty$, and $t \leftarrow 0$.
Solve $\bar{D}(\mu_t(\omega), P_t(\omega), \omega)$ to find $\bar{D}_t$ and $(x_t(\omega), y_t(\omega))$ for all $\omega \in \Omega$.
Compute $\bar{z}(x_t(\omega))$ for some $\omega \in \Omega$ and update z_{UB}
Set $\hat{D}_{t+1} \leftarrow \bar{D}_t$.
Set $\hat{\mu}_{t+1}(\omega) \leftarrow \mu_t(\omega)$ and $\hat{P}_{t+1}(\omega) \leftarrow P_t(\omega)$ for $\omega \in \Omega$.
Add the cuts (26d) at $(x_t(\omega), y_t(\omega))$ for all $\omega \in \Omega$.
loop
  Set $t \leftarrow t + 1$
  Solve the bundle master (26) to find z_t and $(\mu_t(\omega), P_t(\omega))$.
  Solve $\bar{D}(\mu_t(\omega), P_t(\omega), \omega)$ to find $\bar{D}_t$ and $(x_t(\omega), y_t(\omega))$ for all $\omega \in \Omega$.
  Compute $\bar{z}(x_t(\omega))$ for some $\omega \in \Omega$ and update z_{UB}
  if $\min\{z_t, z_{UB}\} - \hat{D}_t \leq \delta(1 + |\hat{D}_t|)$ **then**
    Stop.
  end if
  if $\bar{D}_t \geq \hat{D}_t + \eta(z_t - \hat{D}_t)$ **then**
    Set $\hat{D}_{t+1} \leftarrow \bar{D}_t$.
    Set $\hat{\mu}_{t+1}(\omega) \leftarrow \mu_t(\omega)$ and $\hat{P}_{t+1}(\omega) \leftarrow P_t(\omega)$ for $\omega \in \Omega$.
    Choose $\Delta_{t+1} \in [\Delta_t, \bar{\Delta}]$.
  else
    Set $\hat{D}_{t+1} \leftarrow \hat{D}_t$.
    Set $\hat{\mu}_{t+1}(\omega) \leftarrow \hat{\mu}_t(\omega)$ and $\hat{P}_{t+1}(\omega) \leftarrow \hat{P}_t(\omega)$ for $\omega \in \Omega$.
    Add the cuts (26d) at $(x_t(\omega), y_t(\omega))$ for all $\omega \in \Omega$.
    Choose $\Delta_{t+1} \in [\underline{\Delta}, \Delta_t]$.
  end if
end loop

Theorem 2 (Theorem 2.4 in [18]) *Assume that the subproblems (17) can be solved within a finite time. Algorithm 1 finds a sequence* $\{(\mu_t(\omega), P_t(\omega))\}$ *of dual iterates such that* $\lim_{t\to\infty} \sum_{\omega\in\Omega_t} \hat{D}_t = z_{LD}$.

The convergence property is exactly the same as the BTR method developed in [18]. We remark that by Theorem 2, Algoritmm 1 terminates after a finite number of iterations with δ-optimum for $\delta > 0$. We also remark that the algorithm convergence result is independent of the choice of the TR norm (see Lemma 2.3 in [18]) and the bundle management strategy.

A global optimal solution can be found by employing a branch-and-bound method. For example, a branch-and-bound method has been proposed for branching on the nonanticipativity constraints in [8], which has been improved and implemented by [17]. In particular, the global optimal solutions are reported in [17] for all the SIPLIB test instances [3]. The same branching method, as in [8, 17], can be implemented to the dual decomposition for DRMIP.

6 Computational Experiments

We present our computational experiments for solving DRMIP test instances by using Algorithm 1. The aim of the experiments is to demonstrate (i) the computational performance of the dual decomposition method for discretized

problems and (ii) the impact of the discretization property. As briefly discussed at the end of Section 4, the Lagrangian dual problem (21) suffers from numerical instability with arbitrarily large bounds for α and β_s. Therefore, the numerical results from the problem (21) could not obtained or reported in this section.

6.1 Implementation

We have implemented Algorithm 1 in the open-source software package DSP v1.5.2 [20]. We have used and modified the existing bundle TR method developed in [18] and implemented in DSP. The master problem (26) and scenario subproblems (17) were solved by CPLEX v20.1.0.0. The master problem was solved by using a barrier method without crossover. The choice of barrier method over simplex method has been extensively discussed in the literature (e.g., [28,25,20]) for its implicit regularization effect. However, the simplex method was used only for the iterations at which the barrier method resulted in numerical issue. All the other CPLEX parameters were set to the default values.

We have generated test instances by using the Julia StructJuMP package [15] and writing the StructJuMP models into SMPS file format [7] with one additional file to specify reference scenarios $\hat{\omega}_s$, their probabilities $\hat{p}_s$, and the Wasserstein size ϵ. Hence, DSP reads a DRMIP problem from four files (i.e., *.cor, *.tim, *.sto, and *.dro). All computations were run on a Linux workstation with Intel Xeon Gold 6140 CPU@2.30 GHz and 512 GB of RAM. For a given number k of possible events ω in Ω, we set $\mu_0(\omega) = c/k$, and $P_0(\omega) = \hat{p}_s$ if $\hat{\omega}_s = \omega$; otherwise, $P_0(\omega) = 0$. Note that the initial dual value is feasible to the Lagrangian dual problem (16). We also set the parameters $\underline{\rho} = 3$, $\bar{\rho} = 4$, $\underline{\Delta} = 10^{-2}$, $\bar{\Delta} = 10^{4}$, $\Delta_0 = 0.1$, $\eta = 10^{-4}$, and $\delta = 10^{-5}$. We used ℓ_∞-norm for the TR constraints (26e) and (26f). At each iteration, we choose only one first-stage solution $x_t(\omega)$ to compute an upper bound.

6.2 Test instances

We generate test instances based on the dynamic capacity acquisition and assignment problem [2], also known as DCAP instances available in SIPLIB [3]. The test problem considers the multiperiod capacity expansion for m resources in order to satisfy the requirements of n tasks over T time periods under uncertain scenarios.

Let $x_{it} \geq 0$ be the continuous variables for the capacity acquisition of resource i at time t, and let $u_{it} \in \{0, 1\}$ be the indicator variable corresponding to the acquisition variable x_{it}. For a capacity expansion decision made at the first stage, the second stage assigns resource i to task j at time t for scenario ω, indicated by variable $y_{ijt}(\omega) \in \{0, 1\}$. Let a_{it} and b_{it} respectively be the variable and fixed cost of resource i at time t, and let $c_{ijt}(\omega)$ be the cost of

assigning resource i to task j at time t under scenario ω. We also denote by $d_{jt}(\omega)$ the task j requirement at time t under scenario ω. As a result, the problem is formulated as follows.

$$
\begin{aligned}
\min_{x,u,y(\omega)} \quad & \sum_{i=1}^{m}\sum_{t=1}^{T}(a_{it}x_{it}+b_{it}u_{it}) + \max_{P\in\mathcal{P}}\mathbb{E}_P\left[\sum_{i=0}^{m}\sum_{j=1}^{n}\sum_{t=1}^{T}c_{ijt}(\omega)y_{ijt}(\omega)\right] \\
\text{s.t.} \quad & x_{it}-u_{it}\le 0 \quad \forall i,t, \\
& -\sum_{\tau=1}^{t}x_{i\tau}+\sum_{j=1}^{n}d_{jt}(\omega)y_{ijt}(\omega)\le 0 \quad \forall i,t,\omega, \\
& \sum_{i=0}^{m}y_{ijt}(\omega)=1 \quad \forall j,t,\omega, \\
& x_{it}\ge 0,\ u_{it}\in\{0,1\}, \quad y_{ijt}(\omega)\in\{0,1\}, \quad \forall i,j,t,\omega.
\end{aligned}
$$

Table 1 describes the characteristics of the test problem instances used in our experiments. The problem has mixed-binary variables in the first stage and pure binary variables in the second stage. We denote by $|\xi(\omega)|$ in the table the dimension of the uncertain parameters. The instances are named with the number k of events and the Wasserstein size ϵ. We generated the instances for $k\in\{20,50,100,200,300\}$ and $\epsilon\in\{1,100,500,1000\}$. For all the instances, we used 10 reference scenarios (i.e., $N=10$) with the corresponding probabilities $\hat{p}_s=0.1$ for $s=1,\dots,N$. As a result, we generated 80 problem instances in total. Note that the instances with $\epsilon=0$ are equivalent to SMIP with the reference scenarios and probabilities.

Table 1: Description of test problem instances

					First stage			Second stage		
Name	m	n	T	$\lvert\xi(\omega)\rvert$	Rows	Cols	Ints	Rows	Cols	Ints
dcap233_k_ϵ	2	3	3	36	7	12	6	15	27	27
dcap243_k_ϵ	2	4	3	48	6	12	6	18	36	36
dcap332_k_ϵ	3	3	2	30	6	12	6	12	24	24
dcap342_k_ϵ	3	4	2	40	6	12	6	14	32	32

We randomly generated the deterministic parameter values of a_{it} and b_{it} from $[5,10]$ and $[10,50]$, respectively. We assume that the supports of $c_{ijt}(\omega)$ and $d_{jt}(\omega)$ are defined by $[5,10]$ and $[0.5,1.5]$, respectively, for every $i>0,j,t$. The support of $c_{0jt}(\omega)$ is defined by $[500,1000]$ for every j,t. That is, the values can be anything randomly in the given ranges. The Wasserstein distance was calculated by using two-norm.

6.3 Performance of the dual decomposition

Tables 2 and 3 present the computational results from the dual decomposition method (Algorithm 1) for solving the DRMIP test instances.

Table 2: Numerical results from the dual decomposition solutions for dcap233_k_ϵ and dcap243_k_ϵ instances for $k \in \{20, 50, 100, 200, 300\}$ and $\epsilon \in \{1, 100, 500, 1000\}$.

Name	Iteration	UB	LB	Gap (%)	Time (s)
dcap233_20_1	57	1880.73	1879.50	0.07	12
dcap233_20_100	75	2100.54	2099.32	0.06	18
dcap233_20_500	114	2741.32	2741.05	0.01	17
dcap233_20_1000	103	2741.32	2741.07	0.01	16
dcap233_50_1	68	1675.51	1674.36	0.07	44
dcap233_50_100	117	2029.82	2027.50	0.11	73
dcap233_50_500	153	2816.81	2813.62	0.11	54
dcap233_50_1000	184	2822.60	2814.93	0.27	62
dcap233_100_1	79	1694.39	1693.06	0.08	156
dcap233_100_100	144	2213.38	2211.42	0.09	240
dcap233_100_500	162	3488.44	3482.17	0.18	143
dcap233_100_1000	136	3546.85	3526.07	0.59	91
dcap233_200_1	82	1640.55	1639.03	0.09	667
dcap233_200_100	157	2284.52	2282.64	0.08	987
dcap233_200_500	193	3556.47	3538.24	0.51	425
dcap233_200_1000	179	3586.52	3549.04	1.05	329
dcap233_300_1	87	1644.77	1643.16	0.10	1798
dcap233_300_100	233	2365.09	2363.40	0.07	3225
dcap233_300_500	349	3557.37	3538.38	0.53	2085
dcap233_300_1000	212	3578.18	3549.03	0.81	881
dcap243_20_1	49	3019.25	3017.75	0.05	11
dcap243_20_100	72	3217.67	3216.18	0.05	17
dcap243_20_500	81	4112.12	4109.48	0.06	16
dcap243_20_1000	89	4122.40	4119.18	0.08	16
dcap243_50_1	75	2813.10	2811.59	0.05	52
dcap243_50_100	72	3041.94	3041.71	0.01	45
dcap243_50_500	141	4116.31	4113.31	0.07	72
dcap243_50_1000	124	4122.40	4119.18	0.08	49
dcap243_100_1	79	2698.69	2696.52	0.08	144
dcap243_100_100	86	2969.73	2967.56	0.07	146
dcap243_100_500	184	4419.90	4398.73	0.48	154
dcap243_100_1000	175	4432.84	4406.02	0.60	154
dcap243_200_1	79	2629.55	2627.46	0.08	507
dcap243_200_100	119	2934.67	2932.54	0.07	648
dcap243_200_500	189	4528.29	4504.64	0.52	437
dcap243_200_1000	204	4547.79	4517.71	0.66	434
dcap243_300_1	98	2575.92	2573.77	0.08	1700
dcap243_300_100	152	2912.39	2910.39	0.07	2207
dcap243_300_500	292	4627.86	4623.75	0.09	1624
dcap243_300_1000	296	4772.96	4766.51	0.14	1282

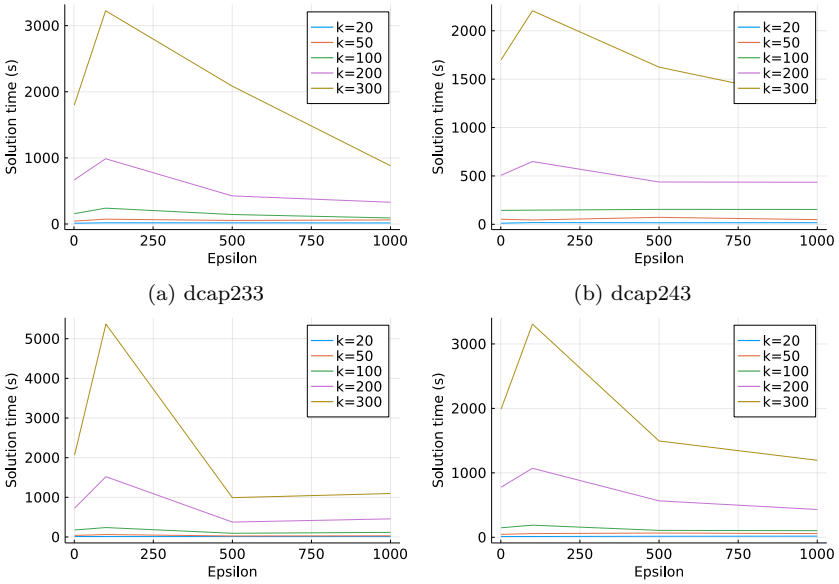
We found the Lagrangian dual bounds (LB) and upper bounds (UB) for all 80 problem instances. Among the 80 problem instances, 72 instances result

Table 3: Numerical results from the dual decomposition solutions for dcap332_k_ϵ and dcap342_k_ϵ instances for $k \in \{20, 50, 100, 200, 300\}$ and $\epsilon \in \{1, 100, 500, 1000\}$.

Name	Iteration	UB	LB	Gap (%)	Time (s)
dcap332_20_1	57	1637.94	1631.04	0.42	12
dcap332_20_100	71	2005.21	1997.67	0.38	11
dcap332_20_500	89	2235.99	2223.13	0.58	11
dcap332_20_1000	83	2236.28	2223.11	0.59	10
dcap332_50_1	65	1649.69	1641.94	0.47	40
dcap332_50_100	130	2366.27	2359.41	0.29	61
dcap332_50_500	101	2910.43	2847.81	2.15	25
dcap332_50_1000	108	2903.17	2847.83	1.91	29
dcap332_100_1	80	1527.75	1518.85	0.58	176
dcap332_100_100	167	2413.72	2406.47	0.30	237
dcap332_100_500	184	3007.50	2951.72	1.85	93
dcap332_100_1000	207	3008.55	2951.70	1.89	116
dcap332_200_1	78	1474.64	1463.33	0.77	725
dcap332_200_100	291	2516.95	2507.81	0.36	1520
dcap332_200_500	231	3169.43	3152.62	0.53	376
dcap332_200_1000	255	3168.99	3152.59	0.52	457
dcap332_300_1	86	1446.34	1435.24	0.77	2069
dcap332_300_100	415	2586.38	2579.03	0.28	5368
dcap332_300_500	272	3165.69	3152.96	0.40	991
dcap332_300_1000	289	3167.63	3152.96	0.46	1096
dcap342_20_1	50	2243.60	2238.53	0.23	11
dcap342_20_100	65	2665.18	2659.50	0.21	13
dcap342_20_500	109	3415.21	3415.16	0.00	16
dcap342_20_1000	124	3418.43	3418.43	0.00	18
dcap342_50_1	67	2052.34	2047.22	0.25	46
dcap342_50_100	104	2522.97	2518.33	0.18	57
dcap342_50_500	190	3418.43	3418.41	0.00	66
dcap342_50_1000	178	3418.43	3418.41	0.00	58
dcap342_100_1	76	1989.02	1981.78	0.36	148
dcap342_100_100	124	2507.82	2500.08	0.31	188
dcap342_100_500	149	3556.09	3549.94	0.17	107
dcap342_100_1000	154	3553.86	3549.93	0.11	103
dcap342_200_1	98	1990.85	1983.75	0.36	777
dcap342_200_100	163	2663.65	2658.36	0.20	1071
dcap342_200_500	264	4173.51	4129.66	1.05	565
dcap342_200_1000	249	4180.99	4144.94	0.86	431
dcap342_300_1	93	1983.80	1976.62	0.36	1989
dcap342_300_100	224	2749.20	2744.43	0.17	3308
dcap342_300_500	322	4192.24	4140.68	1.23	1495
dcap342_300_1000	267	4199.15	4144.96	1.29	1195

in an optimality gap less than 1%, of which 7 instances result in an optimality gap less than 0.01%. The results imply that the Lagrangian dual problem (16) is effective for finding the solutions of DRMIP with very small optimality gaps. The number of iterations ranges from 49 to 415. The total solution time ranges from 10 to 5368 seconds. Note, however, that the solution time could be significantly reduced with parallelization, since the solution times were larger for the instances with larger numbers of scenarios. In particular,

the dual decomposition has been shown to scale strongly with the number of scenarios [20].



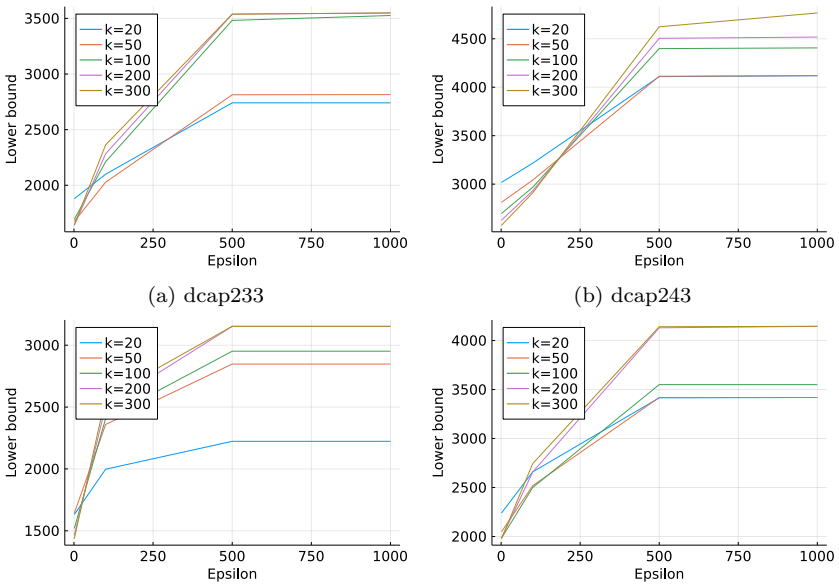
(a) dcap233 (b) dcap243

(c) dcap332 (d) dcap342

Fig. 1: Total solution times for the test instances under different values of the Wasserstein distance limit ϵ for $k \in \{20, 50, 100, 200, 300\}$.

Figure 1 shows the changes of total solution time as the Wasserstein distance limit ϵ increases for each $k \in \{20, 50, 100, 200, 300\}$. Significant increases in the total solution time are observed with the Wasserstein distance limit $\epsilon = 100.0$ for the problem instances with larger numbers of scenarios such as $k = 300$. The reason is that the size of the DRMIP problem (9) increases with the number of discretization points and also results in more Lagrangian dual subproblems (17). Moreover, we observe consistently from all the instances that solving each Lagrangian dual subproblem took longer time for a particular Wasserstein distance limit $\epsilon = 100.0$ than the other values of ϵ.

Figure 2 shows the Lagrangian dual bounds z_{LD} of the test instances by increasing the Wasserstein distance limit ϵ for each $k \in \{20, 50, 100, 200, 300\}$. We recall that the Lagrangian dual bound may increase with larger Wasserstein ambiguity set by ϵ (i.e., being more robust to the uncertainty of our empirical distribution). However, the problem instances with smaller numbers of discretization points k can result in the lower bound being significantly lower than the instances with larger numbers of discretization points, particularly with larger ϵ. This result implies that a sufficiently large number of discretization points are significant to achieve statistically robust solution. We



(a) dcap233 (b) dcap243

(c) dcap332 (d) dcap342

Fig. 2: Lagrangian dual bounds $z_{LD}(\Omega_k)$ for the test instances under different values of the Wasserstein distance limit ϵ for $k \in \{20, 50, 100, 200, 300\}$.

emphasize that the dual decomposition can potentially be parallelized to scale with the number of discretization points, which will allow us to address finer discretizations of Ω without much increase in solution time. In fact, the strong scalability of the method has been shown for SMIP [20,18].

6.4 Numerical comparison with Benders decomposition

Benders decomposition is based on the outer approximation of the second-stage recourse function. The distributionally robust variant has been developed for the problems with pure-binary varialbes in the first stage [5]. Therefore, the distributionally robust Benders decomposition method can also be used by binarizing the first-stage continuous variables. We have implemented the method, as described in [5], in DSP, where the Benders master problem is solved by the SCIP Optimization Suite 7.0.1 [12] and the distribution separation problem and scenario subproblems are solved by CPLEX 12.8.0. The algorithmic differences from [5] include (i) that we use a single cut strategy, as compared with a sophisticated, adaptive strategy, and (ii) that we use the initial lower bound obtained by the dual decomposition.

In this section, we report the numerical results from the distributionally robust Benders method in compariton to those from the dual decomposition. To this end, each first-stage continuous variable x_{it} is binarized with only five

binary varialbes for each x_{it}; that is, $x_{it} := \sum_{l=1}^{5} 2^{-(l-1)} z_{itl}$, where z_{itl} is a binary variable for $l = 1, \ldots, 5$. Note that our choice of binarization uniformly discretizes $[0, 2)$ by the interval of 2^{-4}. Therefore, the binarization results in 36 binary varialbes in the first stage for all instances.

Table 4: Numerical restuls from the Benders decomposition solutions for dcap233_k_ϵ and dcap243_k_ϵ for $k \in \{20, 50, 100, 200, 300\}$ and $\epsilon = \{1, 100, 500, 1000\}$.

Name	Nodes	UB	LB	Gap (%)	Time (s)
dcap233_20_1	159902	2649.3	1848.86	43.29%	7200
dcap233_20_100	159594	2721.01	1848.86	47.17%	7200
dcap233_20_500	146872	2741.63	1848.86	48.29%	7200
dcap233_20_1000	150937	2741.55	1848.86	48.28%	7200
dcap233_50_1	69813	2649.11	1639.65	61.57%	7200
dcap233_50_100	67459	2720.86	1639.65	65.94%	7200
dcap233_50_500	77562	2780.77	1639.65	69.60%	7200
dcap233_50_1000	56890	2820.89	1639.65	72.04%	7200
dcap233_100_1	41421	2649.14	1656.63	59.91%	7200
dcap233_100_100	41191	2776.16	1656.63	67.58%	7200
dcap233_100_500	39815	3260.64	1656.63	96.82%	7200
dcap233_100_1000	37707	3543.69	1656.63	113.91%	7200
dcap233_200_1	20560	2649.17	1598.7	65.71%	7200
dcap233_200_100	20538	2775.88	1598.7	73.63%	7200
dcap233_200_500	17418	3260.74	1598.7	103.96%	7200
dcap233_200_1000	18468	3577.25	1598.7	123.76%	7200
dcap233_300_1	12277	2649.31	1600.96	65.48%	7200
dcap233_300_100	13250	2775.99	1600.96	73.40%	7200
dcap233_300_500	10213	3260.68	1600.96	103.67%	7200
dcap233_300_1000	13507	3577.25	1600.96	123.44%	7200
dcap243_20_1	124133	3839.25	2981.26	28.78%	7200
dcap243_20_100	117127	3981.84	2981.26	33.56%	7200
dcap243_20_500	116424	4122.8	2981.26	38.29%	7200
dcap243_20_1000	122630	4122.8	2981.26	38.29%	7200
dcap243_50_1	58109	3839.23	2764.42	38.88%	7200
dcap243_50_100	54110	3981.86	2764.42	44.04%	7200
dcap243_50_500	50617	4122.84	2764.42	49.14%	7200
dcap243_50_1000	51908	4122.84	2764.42	49.14%	7200
dcap243_100_1	31015	3839.26	2648.54	44.96%	7200
dcap243_100_100	31709	3988.01	2648.54	50.57%	7200
dcap243_100_500	29624	4406.41	2648.54	66.37%	7200
dcap243_100_1000	28636	4432.59	2648.54	67.36%	7200
dcap243_200_1	15329	3839.22	2580.99	48.75%	7200
dcap243_200_100	14905	3988.06	2580.99	54.52%	7200
dcap243_200_500	15806	4414.16	2580.99	71.03%	7200
dcap243_200_1000	15413	4545.29	2580.99	76.11%	7200
dcap243_300_1	10163	3839.25	2527.54	51.90%	7200
dcap243_300_100	9969	3988.11	2527.54	57.79%	7200
dcap243_300_500	9448	4429.69	2527.54	75.26%	7200
dcap243_300_1000	9526	4767.08	2862.44	66.54%	7200

In Tables 4 and 5, the numerical results are reported by using the method for solving the test instances. All the experiments were run with 2-hour wall-

Table 5: Numerical restuls from the Benders decomposition solutions for dcap332_k_ϵ and dcap342_k_ϵ for $k \in \{20, 50, 100, 200, 300\}$ and $\epsilon = \{1, 100, 500, 1000\}$.

Name	Nodes	UB	LB	Gap (%)	Time (s)
dcap332_20_1	256705	2100.05	1531.3	37.14%	7200
dcap332_20_100	227826	2176.15	1531.3	42.11%	7200
dcap332_20_500	210121	2217.05	1531.3	44.78%	7200
dcap332_20_1000	21977	2225.39	1531.3	45.33%	7200
dcap332_50_1	117348	2100.02	1530.29	37.23%	7200
dcap332_50_100	99275	2256.79	1530.29	47.47%	7200
dcap332_50_500	108175	2847.52	1530.29	86.08%	7200
dcap332_50_1000	108554	2897.96	1530.29	89.37%	7200
dcap332_100_1	58351	2100.32	1415.7	48.36%	7200
dcap332_100_100	52103	2267.34	1415.7	60.16%	7200
dcap332_100_500	47723	2902.85	1415.7	105.05%	7200
dcap332_100_1000	55083	3004.84	1415.7	112.25%	7200
dcap332_200_1	24410	2101.06	1358.38	54.67%	7200
dcap332_200_100	26053	2267.44	1358.38	66.92%	7200
dcap332_200_500	20763	2903.23	1358.38	113.73%	7200
dcap332_200_1000	26032	3156.72	1358.38	132.39%	7200
dcap332_300_1	16248	2102.09	1330.77	57.96%	7200
dcap332_300_100	13173	2268.32	1330.77	70.45%	7200
dcap332_300_500	15111	2903.98	1330.77	118.22%	7200
dcap332_300_1000	19315	3159.16	1330.77	137.39%	7200
dcap342_20_1	189862	3169.97	2206.48	43.67%	7200
dcap342_20_100	173728	3284.55	2206.48	48.86%	7200
dcap342_20_500	162907	3402.65	2206.48	54.21%	7200
dcap342_20_1000	174926	3419.19	2206.48	54.96%	7200
dcap342_50_1	84193	3170	1998.52	58.62%	7200
dcap342_50_100	79907	3284.8	1998.52	64.36%	7200
dcap342_50_500	79770	3402.56	1998.52	70.25%	7200
dcap342_50_1000	79805	3419.46	1998.52	71.10%	7200
dcap342_100_1	43322	3170.18	1931.5	64.13%	7200
dcap342_100_100	44374	3284.99	1931.5	70.07%	7200
dcap342_100_500	43637	3484.77	1931.5	80.42%	7200
dcap342_100_1000	41648	3552.56	1931.5	83.93%	7200
dcap342_200_1	18591	3170.3	1932.86	64.02%	7200
dcap342_200_100	21660	3330.21	1932.86	72.29%	7200
dcap342_200_500	21250	3915.8	1932.86	102.59%	7200
dcap342_200_1000	23933	4178.01	1932.86	116.16%	7200
dcap342_300_1	12973	3171.03	1923.77	64.83%	7200
dcap342_300_100	12413	3331.89	1923.77	73.20%	7200
dcap342_300_500	13352	3915.81	1923.77	103.55%	7200
dcap342_300_1000	12962	4198.51	1923.77	118.24%	7200

clock time limit. For all the instances, the Benders method terminated due to the time limit with the optimality gaps ranging from 28% to 137%. The poor performance is mainly due to the weak bounds generated by the integer Benders cuts (e.g., [21]), which eventually leads to an enormous branch-and-bound search tree. This is particularly evident by the following: (i) the number of nodes explored (Nodes) reported in the tables; (ii) that the optimality gap

is almost proportioned to the number of nodes explored; and (iii) that the initial lower bounds were barely improved after the 2-hour solution time.

7 Summary and Directions of Future Work

We have developed the dual decomposition method for solving DRMIP with a finite support. To this end, we first derived the deterministic equivalent form of DRMIP as a linear programming problem. Based on the problem, we presented two different Lagrangian dual problems. One dual problem was derived by applying a Lagrangian relaxation of the nonanticipativity constraints on the first-stage variable and the min-max inequality. The derivation is independent of the description of the ambiguity set. The other dual problem was obtained by using the nonanticipative duality on the first-stage variables and auxiliary variables (α, β_s), which follows the same principle for the dual decomposition of SMIP. We showed that two dual problems result in the same dual bound. However, in our numerical experiments, we found that the latter dual problem suffered from numerical instability if the column bounds are set loosely for (α, β_s). Note that exact column bounds are not given or cannot be obtained in general.

In this paper, we have reported the numerical results from the serial computations of the dual decomposition method for DRMIP. Unlike the dual decomposition of SMIP, the master problem becomes a computational bottleneck due to the incoporation of constraints that represent the Wasserstein ambiguity set. Addressing the challenge will be of interest as future work.

References

1. Ahmed, S.: A scenario decomposition algorithm for 0–1 stochastic programs. Operations Research Letters **41**(6), 565–569 (2013)
2. Ahmed, S., Garcia, R.: Dynamic capacity acquisition and assignment under uncertainty. Annals of Operations Research **124**(1-4), 267–283 (2003)
3. Ahmed, S., Garcia, R., Kong, N., Ntaimo, L., Parija, G., Qiu, F., Sen, S.: SIPLIB: A stochastic integer programming test problem library (2015). URL http://www.isye.gatech.edu/~sahmed/siplib
4. Ahmed, S., Luedtke, J., Song, Y., Xie, W.: Nonanticipative duality, relaxations, and formulations for chance-constrained stochastic programs. Mathematical Programming **162**(1-2), 51–81 (2017)
5. Bansal, M., Huang, K.L., Mehrotra, S.: Decomposition algorithms for two-stage distributionally robust mixed binary programs. SIAM Journal on Optimization **28**(3), 2360–2383 (2018)
6. Bansal, M., Mehrotra, S.: On solving two-stage distributionally robust disjunctive programs with a general ambiguity set. European Journal of Operational Research **279**(2), 296–307 (2019)
7. Birge, J.R., Dempster, M.A., Gassmann, H.I., Gunn, E., King, A.J., Wallace, S.W.: A standard input format for multiperiod stochastic linear programs. Tech. rep., International Institute for Applied Systems Analysis (1987)
8. CarøE, C.C., Schultz, R.: Dual decomposition in stochastic integer programming. Operations Research Letters **24**(1-2), 37–45 (1999)

9. Dandurand, B., Kim, K.: Scalable decomposition methods for preventive security-constrained optimal power flow. In: 2018 Power Systems Computation Conference (PSCC) (2018)
10. Dandurand, B., Kim, K., Schanen, M.: Toward a scalable robust security-constrained optimal power flow using a proximal projection bundle method. Electric Power Systems Research **189**, 106681 (2020)
11. Esfahani, P.M., Kuhn, D.: Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. Mathematical Programming **171**(1-2), 115–166 (2018)
12. Gamrath, G., Anderson, D., Bestuzheva, K., Chen, W.K., Eifler, L., Gasse, M., Gemander, P., Gleixner, A., Gottwald, L., Halbig, K., Hendel, G., Hojny, C., Koch, T., Le Bodic, P., Maher, S.J., Matter, F., Miltenberger, M., Mühmer, E., Müller, B., Pfetsch, M.E., Schlösser, F., Serrano, F., Shinano, Y., Tawfik, C., Vigerske, S., Wegscheider, F., Weninger, D., Witzig, J.: The SCIP Optimization Suite 7.0. Technical report, Optimization Online (2020). URL http://www.optimization-online.org/DB_HTML/2020/03/7705.html
13. Gao, R., Kleywegt, A.J.: Distributionally robust stochastic optimization with wasserstein distance. arXiv preprint arXiv:1604.02199 (2016)
14. Guo, G., Hackebeil, G., Ryan, S.M., Watson, J.P., Woodruff, D.L.: Integration of progressive hedging and dual decomposition in stochastic integer programs. Operations Research Letters **43**(3), 311–316 (2015)
15. Huchette, J., Lubin, M., Petra, C.: Parallel algebraic modeling for stochastic optimization. In: 2014 First Workshop for High Performance Technical Computing in Dynamic Languages, pp. 29–35. IEEE (2014)
16. Kim, K., Anitescu, M., Zavala, V.M.: An asynchronous decomposition algorithm for security constrained unit commitment under contingency events. In: 2018 Power Systems Computation Conference (PSCC) (2018)
17. Kim, K., Dandurand, B.: Scalable branching on dual decomposition of stochastic mixed-integer programming problems. Mathematical Programming Computation **14**, 1–41 (2022)
18. Kim, K., Petra, C.G., Zavala, V.M.: An asynchronous bundle-trust-region method for dual decomposition of stochastic mixed-integer programming. SIAM Journal on Optimization **29**(1), 318–342 (2019)
19. Kim, K., Zavala, V.M.: Large-scale stochastic mixed-integer programming algorithms for power generation scheduling. In: Alternative Energy Sources and Technologies, pp. 493–512. Springer International Publishing (2016)
20. Kim, K., Zavala, V.M.: Algorithmic innovations and software for the dual decomposition method applied to stochastic mixed-integer programs. Mathematical Programming Computation **10**(2), 225–266 (2018)
21. Laporte, G., Louveaux, F.V.: The integer L-shaped method for stochastic integer programs with complete recourse. Operations research letters **13**(3), 133–142 (1993)
22. Linderoth, J., Wright, S.: Decomposition algorithms for stochastic programming on a computational grid. Computational Optimization and Applications **24**(2-3), 207–250 (2003)
23. Lubin, M., Martin, K., Petra, C.G., Sandıkçı, B.: On parallelizing dual decomposition in stochastic integer programming. Operations Research Letters **41**(3), 252–258 (2013)
24. Luo, F., Mehrotra, S.: A decomposition method for distributionally-robust two-stage stochastic mixed-integer cone programs. arXiv preprint arXiv:1911.08713 (2019)
25. Mitchell, J.E.: Computational experience with an interior point cutting plane algorithm. SIAM Journal on Optimization **10**(4), 1212–1227 (2000)
26. Ryan, K., Rajan, D., Ahmed, S.: Scenario decomposition for 0-1 stochastic programs: Improvements and asynchronous implementation. In: 2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp. 722–729. IEEE (2016)
27. Ryu, M., Kim, K.: A privacy-preserving distributed control of optimal power flow. IEEE Transactions on Power Systems **37**(3), 2042–2051 (2021)
28. Sarkissian, J.G.R., Logilab, H.: Column generation with a primal-dual method. Tech. rep., University of Geneva, Switzerland (1996)

29. Sen, S., Sherali, H.D.: Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming. Mathematical Programming **106**(2), 203–223 (2006)
30. Subramanyam, A., Tonbari, M.E., Kim, K.: Data-driven two-stage conic optimization with rare high-impact zero-one uncertainties. arXiv preprint arXiv:2001.04934 (2020)
31. Xie, W.: Tractable reformulations of distributionally robust two-stage stochastic programs with ∞-Wasserstein distance. arXiv preprint arXiv:1908.08454 (2019)
32. Xie, W., Ahmed, S.: Distributionally robust simple integer recourse. Computational Management Science **15**(3-4), 351–367 (2018)
33. Zhang, W., Kim, K., Zavala, V.M.: On the tightness and scalability of the lagrangian dual bound for the alternating current optimal power flow problem. arXiv preprint arXiv:2104.03788 (2021)