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. In addition, we analyze the structural properties showing that DRMIP, as infinite-dimensional programming, is finitely reducible and thus weakly discretizable. We also prove that the Lagrangian dual of DRMIP is weakly discretizable. Based on these properties, we justify the dual decomposition method for solving DRMIP with a discretized support. 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 DCAP test instances (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

This material is based upon work supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357.

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

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

$$\min_{x \in X} \left\{ c^T x + \max_{P \in P} \mathbb{E}_{\xi \sim P} [Q(x, \xi)] \right\}, \quad (1a)$$

where $P$ is an ambiguity set of distributions, $\xi$ is a random variable vector with support $\Xi$, and the second-stage recourse function is defined by

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

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 $d(\xi) \in \mathbb{R}^n_y$, $W(\xi) \in \mathbb{R}^{m_y \times n_y}$, $h(\xi) \in \mathbb{R}^{m_y}$, and $T(\xi) \in \mathbb{R}^{m_y \times n_x}$. 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 a set of distributions around the empirical distribution. To this end, we first define the set of probability measures on support $\Xi$ by

$$\mathcal{M}(\Xi) := \left\{ P \in \mathcal{M} : \int_{\Xi} dP(\xi) = 1 \right\}, \quad (2)$$

where the set $\mathcal{M}$ consists of all probability measures $P : \Xi \to \mathbb{R}_+$. Note that the set of probability measures on any measurable space is convex. Let $\{\xi_1, \ldots, \xi_N\}$ be the set of empirical observations with the corresponding probability estimates $\hat{p}_1, \ldots, \hat{p}_N$. We assume that $\hat{p}_s > 0$ for all $s = 1, \ldots, N$ by removing the empirical observations $\hat{\xi}_s$ with $\hat{p}_s = 0$. We define the ambiguity set of distributions based on the Wasserstein metric at the empirical observations as

$$\mathcal{P} := \left\{ P \in \mathcal{M}(\Xi) : \int_{\Xi} u_s(\xi)d\xi = \hat{p}_s \quad \forall s = 1, \ldots, N, \right\}, \quad (3)$$

where $\epsilon$ is the Wasserstein distance limit. Note that $\| \cdot \|$ may represent any norm.
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., [6,23,7,30]), 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 [6]. A disjunctive programming method, developed for SMIP [28], has been adapted to solve DRMIP within the Benders decomposition framework [23]. However, the class of two-stage DRMIP addressed in [6,23] 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 [9] and has been used successfully for solving SMIP (e.g., [18,1,26,4,13,22]). 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 mixed-integer programming algorithms and solvers. Moreover, an open-source software package DSP [18] enables researchers to prototype, test, and develop new algorithmic features for SMIP. For example, the dual decomposition has been advanced to asynchronous parallel computation [17] and a global optimization solver with branch-and-bound method [16].

In this paper, we develop a dual decomposition method for solving DRMIP under the Wasserstein ambiguity set $\mathcal{P}$. To this end, we derive a deterministic equivalent problem of DRMIP as an infinite-dimensional programming problem. 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 analyze the structural properties of DRMIP and the Lagrangian dual problems. In particular, we show that DRMIP, as an infinite-dimensional program, is finitely reducible and weakly discretizable. In addition, we show that the Lagrangian dual problems are weakly discretizable. These properties suggest that a sufficiently large number of discretization points can effectively approximate both DRMIP and the Lagrangian dual. This is particularly useful and promising with the dual decomposition method, which has been shown to strongly scale with the number of scenarios in SMIP [18,17]. 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 2 we present the deterministic equivalent formulation of DRMIP. Section 3 presents the Lagrangian dual problems. In Section 4 we show the discretization properties for DRMIP and the Lagrangian dual. 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 Deterministic Equivalent Formulation

We present the deterministic equivalent formulation of DRMIP, which is of infinite-dimensional mixed-integer programming for a continuous support Ξ. Note that the formulation becomes finite-dimensional for a finite support. The following duality property is key to deriving our reformulation. We assume that P is nonempty with N > 1 throughout this paper.

**Lemma 1** For any random variable f(ξ) ∈ R, the strong duality property holds for the following problem:

\[
\max_{P \in \mathcal{P}} \mathbb{E}_P[f(\xi)].
\] (4)

Furthermore, its dual is given as the following semi-infinite program:

\[
\min_{\alpha \geq 0, \beta_s} \epsilon \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s
\] (5a)

subject to

\[
\| \hat{\xi}_s - \xi \| \alpha + \beta_s \geq f(\xi) \quad \forall \xi \in \Xi, \ s = 1, \ldots, N.
\] (5b)

**Proof** We observe that the maximization problem is an infinite-dimensional linear programming problem that consists of infinitely many variables and constraints. Note that such a problem has nontrivial duality. By eliminating \( P(\xi) \), the infinite-dimensional linear program can be reformulated as the following semi-infinite linear programming form,

\[
\max_{u_s(\xi) \geq 0} \int_{\Xi} \sum_{s=1}^{N} f(\xi) u_s(\xi) d\xi
\] (6a)

subject to

\[
\int_{\Xi} \sum_{s=1}^{N} u_s(\xi) \| \hat{\xi}_s - \xi \| d\xi \leq \epsilon, \quad (\alpha)
\] (6b)

\[
\int_{\Xi} u_s(\xi) d\xi = \hat{p}_s \quad \forall s = 1, \ldots, N, \quad (\beta_s)
\] (6c)

\[
\int_{\Xi} \sum_{s=1}^{N} u_s(\xi) d\xi = 1.
\] (6d)
Note that constraint (6d) is implied by constraints (6c) and can be removed from the problem (6). Let \((\alpha, \beta_s)\) be the Lagrangian multipliers with respect to the constraints. Then, the dual of (6) is given in (5).

Note that the weak duality holds. By Theorem 2.1 in [12], the linear semi-infinite system of (5) is consistent because the system is feasible at \((\alpha, \beta_s) = (0, \max_{\xi \in \Xi} f(\xi), \ldots, \max_{\xi \in \Xi} f(\xi))\). In addition, since \(\hat{p}_s > 0\) for all \(s = 1, \ldots, N\) and \(N > 1\), the objective coefficient vector \([\epsilon, \hat{p}_1, \ldots, \hat{p}_N]\) of problem (5) is in the relative interior of the convex cone

\[
\text{cone} \left\{ \left( \|\hat{\xi}_s - \xi\| e_s \right) \mid \forall \xi \in \Xi, s = 1, \ldots, N \right\},
\]

where \(e_s\) is the \(s\)th column of the \(N \times N\)-identity matrix. Therefore, by Theorem 4(v) in [12], the problem (5) has the duality gap of zero. \(\Box\)

Although the strong duality shown in Lemma 1 is trivial for a finite support \(\Xi\), the duality gap can be positive for a linear infinite-dimensional program as well as a linear semi-infinite program (e.g., [14]). Moreover, the strong duality result allows us to derive the reformulation of two-stage DRMIP with a general set of \(\Xi\).

Proposition 1 Suppose that \(P\) is defined by (3). Two-stage DRMIP (1) can be written as a deterministic infinite-dimensional program of the form

\[
z(\Xi) = \min_{x, \alpha, \beta_s, y(\xi)} \ c^T x + \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s \quad (7a)
\]

\[
s.t. \qquad \|\hat{\xi}_s - \xi\| \alpha + \beta_s - d(\xi)^T y(\xi) \geq 0 \quad \forall \xi \in \Xi, s = 1, \ldots, N, \quad (7b)
\]

\[
T(\xi)x + W(\xi)y(\xi) \geq h(\xi) \quad \forall \xi \in \Xi, \quad (7c)
\]

\[
x \in X, \quad \alpha \geq 0, \quad y(\xi) \in Y \quad \forall \xi \in \Xi. \quad (7d)
\]

Proof The problem (1) can be written as the following semi-infinite program:

\[
\min_{x \in X, r \in \mathbb{R}, y(\xi) \in Y} \ c^T x + r \quad (8a)
\]

\[
s.t. \quad r \geq \mathbb{E}_P \left[ d(\xi)^T y(\xi) \right] \quad \forall P \in \mathcal{P}, \quad (8b)
\]

\[
T(\xi)x + W(\xi)y(\xi) \geq h(\xi) \quad \forall \xi \in \Xi. \quad (8c)
\]

The constraint (8b) is equivalent to \(r \geq \max_{P \in \mathcal{P}} \mathbb{E}_P \left[ d(\xi)^T y(\xi) \right] = (5)\) by Lemma 1, and thus

\[
r \geq \epsilon \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s \quad (9a)
\]

\[
\|\hat{\xi}_s - \xi\| \alpha + \beta_s \geq d(\xi)^T y(\xi) \quad \forall \xi \in \Xi, s = 1, \ldots, N. \quad (9b)
\]

Substituting variable \(r\) in (7a) and (9a) results in (7). \(\Box\)
The infinite-dimensional programming reformulation (7) embeds the block-angular structure that is observed in SMIP, in which variables \((x, \alpha, \beta_s, \gamma)\) can be considered as the first-stage (i.e., nonanticipative) variables. As a result, the existing theories and algorithms for SMIP can be applied to the distributionally robust variant (7). Most important, the convergence properties of the sample average approximation (SAA) for SMIP \([5,19]\) hold for the distributionally robust variant.

We conclude this section by presenting the convergence result of objective values and solutions in \([19]\). Let \(\{\xi_1, \ldots, \xi_k\} =: \Xi_k \subset \Xi\) be an independently and identically distributed (i.i.d.) random sample of \(k\) realizations. For a given \(\Xi_k\), we define the following SAA of DRMIP (7) as

\[
\begin{align*}
  z(\Xi_k) &= \min_{x, \alpha, \beta_s, y(\xi)} c^T x + \epsilon \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s \\
  \text{s.t.} & \quad \|\hat{\xi}_s - \xi\| \alpha + \beta_s - d(\xi)^T y(\xi) \geq 0 \quad \forall \xi \in \Xi_k, \ s = 1, \ldots, N, \\
                & \quad T(\xi) x + W(\xi)y(\xi) \geq h(\xi) \quad \forall \xi \in \Xi_k, \\
                & \quad x \in X, \ \alpha \geq 0, \ \ y(\xi) \in Y \quad \forall \xi \in \Xi_k.
\end{align*}
\]

Moreover, let \(S\) and \(S_k\) be the sets of optimal solutions for problems (7) and (10), respectively.

**Proposition 2 (Proposition 2.1 in [19])** The following two properties hold: (i) \(z(\Xi_k) \to z(\Xi)\) w.p.1 as \(k \to \infty\), and (ii) the event \(\{S_k \subset S\}\) happens w.p.1. for \(k\) large enough.

### 3 Lagrangian Dual of DRMIP

It may be impractical to solve the deterministic equivalent problem even for a discretization \(\Xi_k\) of the support, particularly with large \(k\). In this section, we present the Lagrangian dual of DRMIP, which would enable us to use the existing dual decomposition methods (e.g., \([17,22,16,26]\)) 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 (3), by applying the Lagrangian relaxation technique used for SMIP. In particular, the problem (7) can be seen as a SMIP, where \((x, \alpha, \beta_s)\) are the first-stage variables and \(y(\xi)\) 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 3 The Lagrangian dual of problem (1) is given by

\[
\begin{align*}
  z_{LD}(\Xi) &= \max_{\lambda(\xi), P \in \mathcal{P}} \mathbb{E}_P [D(\lambda(\xi), \xi)] \\
  \text{s.t.} \quad &\mathbb{E}_P [\lambda(\xi)] = c,
\end{align*}
\]

where

\[
\begin{align*}
  D(\lambda(\xi), \xi) &= \min_{x \in X, y \in Y} \lambda(\xi)^T x + d(\xi)^T y(\xi) \\
  \text{s.t.} \quad &T(\xi)x + W(\xi)y \geq h(\xi).
\end{align*}
\]

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

\[
\begin{align*}
  \min_{x, x(\xi) \in X} &\left\{ c^T x + \max_{P \in \mathcal{P}} \mathbb{E}_P [Q(x(\xi), \xi)] \right\} \\
  \text{s.t.} \quad &x - x(\xi) = 0 \quad \forall \xi \in \Xi, \\
  &x(\xi) \in X \quad \forall \xi \in \Xi,
\end{align*}
\]

where \((13b)\) are called the nonanticipativity constraints. By taking the Lagrangian relaxation with respect to \((13b)\), we have the following Lagrangian function:

\[
\begin{align*}
  \min_{x, x(\xi) \in \Xi} &\left\{ c - \int_{\Xi} \mu(\xi) d\xi \right\}^T x + \int_{\Xi} \mu(\xi)^T x(\xi) d\xi + \max_{P \in \mathcal{P}} \mathbb{E}_P [Q(x(\xi), \xi)]
\end{align*}
\]

where \(\mu(\xi)\) are the Lagrangian multipliers corresponding to the constraints \((13b)\). For the Lagrangian function value to be bounded, we need

\[
\int_{\Xi} \mu(\xi) d\xi = c,
\]

and thus the corresponding term in the objective function is eliminated. The objective function can be written as

\[
\max_{P \in \mathcal{P}} \int_{\Xi} P(\xi) \left[ \frac{\mu(\xi)}{P(\xi)} x(\xi) + Q(x(\xi), \xi) \right] d\xi = \max_{P \in \mathcal{P}} \mathbb{E}_P [\lambda(\xi)x(\xi) + Q(x(\xi), \xi)],
\]

where \(\mu(\xi) = \lambda(\xi)P(\xi)\). In addition, by the min-max inequality, we have

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

By substituting \(\mu(\xi) = \lambda(\xi)P(\xi)\) in \((14)\), we have the dual problem \((11)\). \(\Box\)

Note that the weak duality holds; that is, \(z(\Xi) \geq z_{LD}(\Xi)\). The Lagrangian dual problem \((11)\) can be seen as a direct extension from SMIP to DRMIP and is equivalent to that of SMIP for a singleton \(P\) (i.e., when the distribution is known). As in the dual decomposition of SMIP, evaluating the Lagrangian dual function \(D\) can be parallelized for a finite \(\Xi\). In addition, regardless of the specification of \(P\), the scenario subproblems \(D(\lambda(\xi), \xi)\) are equivalent to those
for SMIP. The Lagrangian dual (11) of DRMIP differs from that of SMIP only by perturbing the probability measure defined in $P$.

However, the Lagrangian dual problem (11) 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 (11) is equivalent to

$$z_{LD}(\Xi) = \max_{\mu(\xi), P \in \mathcal{P}} \int_{\xi} \hat{D}(\mu(\xi), P(\xi)) d\xi$$  \hspace{1cm} (15a)  

subject to \hspace{1cm} \int_{\Xi} \mu(\xi) d\xi = c,  \hspace{1cm} (15b)$$

where

$$\hat{D}(\mu(\xi), P(\xi), \xi) := \min_{x \in X, y \in Y} \mu(\xi)^T x + P(\xi)d(\xi)^T y$$  \hspace{1cm} (16a)  

subject to \hspace{1cm} T(\xi)x + W(\xi)y \geq h(\xi).  \hspace{1cm} (16b)$$

Proof The equivalent formulation (15) is obtained by substituting $\lambda(\xi)$ by $\mu(\xi)/P(\xi)$ in (11) and (12). \hfill \Box

We now present the Lagrangian dual of the deterministic equivalent DRMIP (7) with the Wasserstein ambiguity set. To this end, we consider the following equivalent formulation of DRMIP (7):

$$\min_{x, \alpha, \beta, x(\xi), \alpha(\xi), \beta(\xi), y(\xi)} c^T x + \epsilon\alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s$$  \hspace{1cm} (17a)  

subject to \hspace{1cm} x - x(\xi) = 0 \hspace{0.5cm} \forall \xi \in \Xi,  \hspace{1cm} (17b)  

$$\alpha - \alpha(\xi) \geq 0 \hspace{0.5cm} \forall \xi \in \Xi,  \hspace{1cm} (17c)$$  

$$\beta_s - \beta_s(\xi) \geq 0 \hspace{0.5cm} \forall \xi \in \Xi, \hspace{0.5cm} s = 1, \ldots, N,  \hspace{1cm} (17d)$$  

$$\|\hat{\xi}_s - \xi\|_2 \alpha(\xi) + \beta_s(\xi) - d(\xi)^T y(\xi) \geq 0$$  \hspace{1cm} \forall \xi \in \Xi, \hspace{0.5cm} s = 1, \ldots, N, \hspace{1cm} (17e)$$  

$$T(\xi)x(\xi) + W(\xi)y(\xi) \geq h(\xi) \hspace{0.5cm} \forall \xi \in \Xi,  \hspace{1cm} (17f)$$  

$$x(\xi) \in X, \hspace{0.5cm} \alpha(\xi) \geq 0, \hspace{0.5cm} y(\xi) \in Y \hspace{0.5cm} \forall \xi \in \Xi,  \hspace{1cm} (17g)$$

where (17b)–(17d) are the nonanticipativity constraints. Note that the reformulation technique used in (17) has been commonly used to derive the dual or scenario decomposition of SMIP (e.g., [9,1,18]).

Proposition 4 The Lagrangian relaxation of (17) with respect to the nonanticipativity constraints (17b), (17c), and (17d) gives the following Lagrangian
dual problem:

\[
\begin{align*}
\text{maximize} & \quad f(\nu, \xi) \\
\text{s.t.} & \quad g(\xi) \\
\end{align*}
\]

where \( f(\nu, \xi) := (\zeta^T \nu + \xi) \) and \( g(\xi) \) are functions of \( \xi \).
which can be decomposed for each $\xi \in \Xi$. With the boundedness conditions (20), we have the Lagrangian dual problem as in (18).

We note that the Lagrangian dual problem (18) may be computationally more expensive than the problem (11). Specifically, each scenario subproblem (19) of the new dual problem (18) has $N + 1$ variables and $N$ constraints more than those of (11). Nevertheless, the Lagrangian dual problem (18) results in the same Lagrangian bound as the problem (11) does.

**Theorem 1** Suppose that $\mathcal{P}$ is defined as (3). $z_{\text{WLD}}(\Xi) = z_{\text{LD}}(\Xi)$.

**Proof** Introducing variables $r_s$, the scenario subproblem (19) for a given $\xi \in \Xi$ is equivalent to

$$\min_{x, \alpha, \beta_s, y} \mu(\xi)^T x + \nu(\xi) \alpha + \sum_{s=1}^{N} u_s(\xi) \beta_s$$

s.t. $\|\hat{\xi}_s - \xi\| \alpha + \beta_s - d(\xi)^T y - r_s = 0 \quad \forall s = 1, \ldots, N$,

(19c) - (19d), $r_s \geq 0 \quad \forall s = 1, \ldots, N$.

Eliminating variables $\beta_s$ from this equivalent problem, we have

$$\min_{x, \alpha, \beta_s, r_s, y} \mu(\xi)^T x + \left[ \nu(\xi) - \sum_{s=1}^{N} u_s(\xi) \|\hat{\xi}_s - \xi\| \right] \alpha$$

$$+ \sum_{s=1}^{N} u_s(\xi) d(\xi)^T y + \sum_{s=1}^{N} u_s(\xi) r_s$$

s.t. (19c) - (19d), $r_s \geq 0 \quad \forall s = 1, \ldots, N$.

We observe that $\alpha = 0$ if

$$\nu(\xi) - \sum_{s=1}^{N} u_s(\xi) \|\hat{\xi}_s - \xi\| \geq 0.$$  (21)

Otherwise, the subproblem is unbounded. In addition, $r_s = 0$ since $u_s(\xi) \geq 0$ for all $s = 1, \ldots, N$. Therefore, the scenario subproblem (19) is equivalent to

$$\min_{x, y} \left\{ \mu(\xi)^T x + \sum_{s=1}^{N} u_s(\xi) d(\xi)^T y : (19c) - (19d) \right\},$$

and thus also equivalent to the scenario subproblem (16) by substituting $\sum_{s=1}^{N} u_s(\xi)$ with $P(\xi)$; that is,

$$P(\xi) = \sum_{s=1}^{N} u_s(\xi) \quad \forall \xi \in \Xi.$$  (22)

By adding the constraints (21) and (22) to the problem (18), we have the Lagrangian dual problem (15).
A key advantage of using the Lagrangian dual problem (18) may be that existing scenario decomposition methods and solvers for SMIP can be used without any modification (e.g. [1,26,16]). Note, however, that scenario sub-problems can be unbounded for some dual variable values, particularly if inequality (21) is violated. The unbounded solutions may be avoided by introducing arbitrarily large bounds to $\alpha$ and $\beta$. 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 $\alpha$ and $\beta$. On the other hand, only minor modification to existing SMIP solvers would be required for solving the Lagrangian dual problem (15).

4 Discretization Properties

In this section, we show that DRMIP (7) is finitely discretizable and reducible, which are the properties borrowed from semi-infinite programming. The key idea for showing these properties is based on the fact that the strong duality property is closely related to the discretization of the dual problem (5), as discussed in [29]. We emphasize, however, that the properties are not trivial to show for the infinite-dimensional programming (7).

Let $v$ be the optimal value of the linear semi-infinite programming (5). For a given finite set $\{\xi_1, \ldots, \xi_k\} =: \Xi_k \subseteq \Xi$, we define the following (finite) linear programming problem:

$$v_k := \min_{\alpha \geq 0, \beta_s} \; \epsilon \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s$$

s.t. $\|\hat{\xi}_s - \xi\|_{\alpha + \beta_s} \geq f(\xi) \quad \forall \xi \in \Xi_k, \; s = 1, \ldots, N$. (23b)

We note that $v \geq v_k$, since the discretization (23) is the relaxation of (5) with respect to the constraints. We now define the discretizable and reducible properties.

**Definition 1** Problem (5) is said to be finitely reducible if there exists a finite discretization $\Xi_k$ such that $v = v_k$.

**Definition 2** Problem (5) is said to be weakly discretizable if there exists a sequence $\{\Xi_k\}$ of finite discretizations such that $v_k \rightarrow v$.

Specifically, the strong duality property in Lemma 1 immediately leads to the following result.

**Theorem 2** Problem (5) is finitely reducible and thus weakly discretizable.

**Proof** The strong duality of problem (5) follows from Lemma 1. Since $P$ is nonempty, the dual of (5) is solvable (i.e., feasible). Therefore, by Theorem 7 in [21], the problem (5) is finitely reducible and thus weakly discretizable. □
We now show that these properties hold for DRMIP (7). Note that the weakly discretizable and reducible properties are not immediate for infinite-dimensional programs such as (7). Let $z$ be the optimal value of the infinite-dimensional mixed-integer programming (7). For a given finite set $\Xi_k \subset \Xi$, we define the following mixed-integer programming problem:

$$
\begin{align*}
  z_k := \min_{x, \alpha, \beta, y(\xi)} & \quad c^T x + \epsilon \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s \\
  \text{s.t.} & \quad \|\hat{\xi}_s - \xi\| \alpha + \beta_s - d(\xi)^T y(\xi) \geq 0 \quad \forall \xi \in \Xi_k, \ s = 1, \ldots, N, \\
  & \quad T(\xi)x + W(\xi) y(\xi) \geq h(\xi) \quad \forall \xi \in \Xi_k, \\
  & \quad x \in X, \ \alpha \geq 0, \ y(\xi) \in Y \quad \forall \xi \in \Xi_k.
\end{align*}
$$

**Theorem 3** DRMIP (7) is finitely reducible and thus weakly discretizable.

**Proof** For any $\Xi_k \subset \Xi$, Lemma 1 and Proposition 1 allow the deterministic equivalent formulation

$$
\begin{align*}
  z(\Xi_k) := \min_{x, \alpha, \beta, y(\xi)} & \quad c^T x + \epsilon \alpha + \sum_{s=1}^{N} \hat{p}_s \beta_s \\
  \text{s.t.} & \quad \|\hat{\xi}_s - \xi\| \alpha + \beta_s - Q(x, \xi) \geq 0 \quad \forall \xi \in \Xi_k, \ s = 1, \ldots, N, \\
  & \quad x \in X, \ \alpha \geq 0.
\end{align*}
$$

For any $\Xi_k \subset \Xi_{k+1}$, we observe that $z(\Xi_{k+1}) \geq z(\Xi_k)$.

Suppose that $\Xi_k$ is a finite discretization of $\Xi$ such that $v = v_k$. Let $P_k$ be the Wasserstein ambiguity set (3) defined for the discrete set $\Xi_k$; that is,

$$
P_k := \left\{ P \in \mathcal{M}(\Xi_k) : \begin{array}{l}
  \sum_{\xi \in \Xi_k} \sum_{s=1}^{N} u_s(\xi) \|\hat{\xi}_s - \xi\| \leq \epsilon, \\
  \sum_{\xi \in \Xi_k} u_s(\xi) = \hat{p}_s \quad \forall s = 1, \ldots, N, \\
  \sum_{s=1}^{N} u_s(\xi) = P(\xi) \quad \forall \xi \in \Xi_k, \\
  u_s(\xi) \geq 0 \quad \forall \xi \in \Xi_k, \ s = 1, \ldots, N
\end{array} \right\}.
$$

For such $\Xi_k$, the problem (25) becomes equivalent to the DRMIP (1) (by Theorem 2) and also equivalent to (24). Consequently, DRMIP is finitely reducible. By taking a subsequence of finite discretization to the reduced discretization $\Xi_k$, we obtain the weak discretizability.

This provides a different view of the convergence of DRMIP with respect to the support, as compared with that of SMIP as shown in Proposition 2. For a continuous support $\Xi$, there may not exist a discretization $\Xi_k$ such that SMIP
is finitely reducible. In other words, it is not guaranteed that a single scenario can represent the entire support for SMIP. The finite reducibility of DRMIP also implies the existence of the worst-case scenario that can represent DRMIP. Moreover, the objective function value of SMIP may not be nondecreasing for $\Xi_k \subseteq \Xi_{k+1}$, whereas that of DRMIP is, as follows.

**Corollary 2** $z(\Xi_{k+1}) \geq z(\Xi_k)$ for any $\Xi_k \subseteq \Xi_{k+1} \subset \Xi$.

Although we derived the Lagrangian dual problems with generic support $\Xi$, a practical computation may require a discretization of the support, as seen in Section 4. By weak duality and Theorem 1, respectively, we have

$$z(\Xi_k) \geq z_{LD}(\Xi_k) = z_{WLD}(\Xi_k)$$

for any discretization $\Xi_k \subset \Xi$. In addition, we need to ensure that the Lagrangian dual problems are also weakly discretizable. However, we note that the question to whether the Lagrangian dual problem (18) is finitely reducible remains open.

**Theorem 4** The Lagrangian dual problem (18) is weakly discretizable.

**Proof** The proof is based on the idea of Dantzig-Wolfe decomposition [10]. We consider the dual of the Lagrangian dual problem (18), which is of the Dantzig-Wolfe decomposition form (e.g., [16]) given by

$$z_{LD}(\Xi) = \min_{x, \alpha, \beta, y(\xi)} c^T x + \epsilon \alpha + \sum_{s=1}^N \hat{p}_s \beta_s$$

s.t. $(x, \alpha, \beta, y(\xi)) \in \text{conv}(G(\xi)) \forall \xi \in \Xi$, \hspace{1cm} (27a)

where $\beta := (\beta_1, \ldots, \beta_N)$, $\text{conv}(G(\xi))$ represents the convex hull of $G(\xi)$ and

$$G(\xi) := \{(x(\xi), a(\xi), \beta(\xi), y(\xi)) : (17e) - (17g)\}.$$.

For any discretization $\Xi_k \subseteq \Xi_{k+1}$, we have $z_{LD}(\Xi_{k+1}) \geq z_{LD}(\Xi_k)$. Moreover, we define $\Xi_k \subseteq \Xi_{k+1} \subset \cdots \subseteq \Xi$ such that $\Xi = \bigcup_{k=1}^{\infty} \Xi_k$. Then, we have $z_{LD}(\Xi) = \lim_{k \to \infty} z_{LD}(\Xi_k)$. \hfill $\Box$

Using the properties of weak discretizability and reducibility, discretization methods can be developed in order to iteratively approximate the DRMIP problem. For example, a discretization method may be implemented at each iteration $k$ as follows:

1. Compute an optimal solution $(x_k, \alpha_k, \beta_{k,s})$ of (25) for a given $\Xi_k \subset \Xi$.
2. Stop if there exists $\xi \in \Xi$ such that $\|\hat{\xi}_s - \xi\| \alpha_k + \beta_{k,s} - Q(x_k, \xi) \geq 0$.
   Otherwise, find a discretization $\Xi_{k+1} \supset \Xi_k$.

More algorithmic approaches are available in [25,14]. Note, however, developing discretization methods is beyond the scope of this paper.
5 Dual Decomposition

We present the dual decomposition method that can be applied to the Lagrangian dual problem (15). We consider a discretization of the Lagrangian dual problem (15) for given $\Xi_k \subset \Xi$ and adapt the dual decomposition method that has been developed for SMIP in [9,18] with the additional dual constraints defined by $P$.

A main idea of the dual decomposition is the outer approximation of the Lagrangian dual function $\bar{D}(\mu(\xi), P(\xi), \xi)$ 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. Extensive numerical experiments are also available in [18]. The bundle methods are also capable of easily incorporating additional constraints from the ambiguity set $P$ as well as regularization (e.g., [18,17]).

After $t$ iterations, the outer approximation of the Lagrangian dual problem (15) is given by

$$z_t(\Xi_k) := \max_{\theta(\xi), \mu(\xi), P(\xi)} \sum_{\xi \in \Xi_k} \theta(\xi)$$

s.t.

$$\sum_{\xi \in \Xi_k} \mu(\xi) = c,$$

$$\langle P(\xi) \rangle_{\xi \in \Xi_k} \in P_k,$$

$$\theta(\xi) \leq \hat{x}_j(\xi)^T \mu(\xi) + d(\xi)^T \hat{y}_j(\xi) P(\xi) \quad \forall j = 1, \ldots, t, \xi \in \Xi_k,$$

$$\|\mu(\xi) - \hat{\mu}_t(\xi)\| \leq \Delta_t \quad \forall \xi \in \Xi_k,$$

$$\|P(\xi) - \hat{P}_t(\xi)\| \leq \Delta_t \quad \forall \xi \in \Xi_k,$$

where $P_k$ is defined as in (26); the right-hand side of constraints (28d) are the subgradients of $\bar{D}$ at each iteration $j = 1, \ldots, t$ for each $\xi \in \Xi_k$; and constraints (28e) and (28f) are the trust-region (TR) constraints with center $(\hat{\mu}_t(\xi), \hat{P}_t(\xi))$ and size $\Delta_t$ at iteration $t$.

We dynamically adjust the TR center and size in order to regularize the iterates. Let $\underline{\Delta}$ and $\overline{\Delta}$ be the minimum and maximum values of the TR size $\Delta_t$, respectively. We denote by $(\hat{x}_t(\xi), \hat{y}_t(\xi))$ an optimal solution of the Lagrangian subproblem (16) for $(\mu_t(\xi), P_t(\xi))$ at iteration $t$. We also denote by $\bar{D}_t := \sum_{\xi \in \Xi_k} \bar{D}(\mu_t(\xi), P_t(\xi), \xi)$ the Lagrangian dual value at iteration $t$. We update the TR center if the Lagrangian function value is sufficiently decreased at iteration $t$ as

$$\hat{D}_t \geq \bar{D}_t + \eta \zeta_t,$$

where $\eta \in (0,0.5)$ and $\zeta_t := z_t(\Xi_t) - \bar{D}_t$ (i.e., predicted increase). Then, the bundle master problem (28) 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 (28d) may be added to improve the outer approximation.

A careful update of the TR size $\Delta_t$ 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 [17] for the TR size $\Delta_t$. The TR size is updated as follows:

1. If $\rho > 0$, then set $\tau_t \leftarrow \tau_t + 1$.
2. If $\rho > \rho_0$ or ($\rho \in (0, \rho]$ and $\tau_t \geq \bar{\rho}$), then set $\tau_t \leftarrow 0$ and $\Delta_t \leftarrow \max\{\Delta_t \min\{\rho, \rho\}, \bar{\Delta}\}$.

Here $\tau_t$ counts the number of iterations in which TR is not reduced, and $\rho$ and $\bar{\rho}$ are given parameters. The approximation error is measured by

$$\rho := \min \left\{ \frac{1}{\tau_t} \left\| \frac{\mu_t(\xi) - \hat{\mu}_t(\xi)}{\tau} \right\|, \frac{\max\{\bar{D}_t - \hat{D}_t, l_t\}}{\zeta_t} \right\},$$

and the model linearization error is

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

On the other hand, we may detect that the TR is too small if the solution is bounded by the TR and if $\hat{D}_t \geq \hat{D}_t + 0.5\zeta_t$. Then, we increase the TR by $\Delta_t \leftarrow \min\{2\Delta_t, \bar{D}\}$. More details of the algorithm development (e.g., serious and null steps, TR updates) are available in [20,17].

We have implemented an algorithm to find upper bounds for first-stage solution $x_t(\xi)$. Suppose that the first-stage solutions $x_t(\xi)$ are given for $\xi \in \Xi_k$ at iteration $t$. We evaluate the recourse function $Q(x_t(\xi), \xi)$ for each $\xi \in \Xi_k$. 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 [18]. Then, an upper bound can be computed at the first-stage solution $x_t$ by solving the following linear programming problem:

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

s.t. $\|\xi_s - \xi\| \alpha + \beta_s \geq Q(x_t, \xi)$ $\forall \xi \in \Xi_k$, $s = 1, \ldots, N$.  

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

Algorithm 1 is adapted from the BTR method [17] that implements a bundle TR method for SMIP. The algorithm initializes a dual feasible solution
Algorithm 1 A dual decomposition method for DRMIP

1: Initialize a feasible \((\mu_0(\xi), P_0(\xi))\) for all \(\xi \in \Xi_k\), \(\Delta_0 \in [\Delta, \bar{\Delta}]\), \(\delta \geq 0\), and \(\eta \in (0, 0.5)\), \(z_{UB} \leftarrow \infty\), and \(t \leftarrow 0\).
2: Solve \(\hat{D}(\mu_1(\xi), P_1(\xi), \xi)\) to find \(\hat{D}_t\) and \((x_t(\xi), y_t(\xi))\) for all \(\xi \in \Xi_k\).
3: Compute \(\hat{z}(x_t(\xi))\) for some \(\xi \in \Xi_k\) and update \(z_{UB}\)
4: Set \(\Delta_{t+1} \leftarrow \hat{D}_t\).
5: Set \(\mu_{t+1}(\xi) \leftarrow \mu_t(\xi)\) and \(P_{t+1}(\xi) \leftarrow P_t(\xi)\) for \(\xi \in \Xi_k\).
6: Add the cuts (28d) at \((x_t(\xi), y_t(\xi))\) for all \(\xi \in \Xi_k\).
7: loop
8: Set \(t \leftarrow t + 1\)
9: Solve the bundle master (28) to find \(z_t(\Xi_k)\) and \((\mu_t(\xi), P_t(\xi))\).
10: Solve \(\hat{D}(\mu_t(\xi), P_t(\xi), \xi)\) to find \(\hat{D}_t\) and \((x_t(\xi), y_t(\xi))\) for all \(\xi \in \Xi_k\).
11: Compute \(\hat{z}(x_t(\xi))\) for some \(\xi \in \Xi_k\) and update \(z_{UB}\)
12: if \(z_t(\Xi_k) - \hat{D}_t \leq \delta(t + |D_t|)\) then
13: Stop.
14: end if
15: if \(\hat{D}_t \geq D_t + \eta(z_t(\Xi_k) - \hat{D}_t)\) then
16: Set \(\hat{D}_{t+1} \leftarrow \hat{D}_t\).
17: Set \(\mu_{t+1}(\xi) \leftarrow \mu_t(\xi)\) and \(P_{t+1}(\xi) \leftarrow P_t(\xi)\) for \(\xi \in \Xi_k\).
18: Choose \(\Delta_{t+1} \in [\Delta_t, \bar{\Delta}]\).
19: else
20: Set \(\hat{D}_{t+1} \leftarrow \hat{D}_t\).
21: Set \(\mu_{t+1}(\xi) \leftarrow \mu_t(\xi)\) and \(P_{t+1}(\xi) \leftarrow P_t(\xi)\) for \(\xi \in \Xi_k\).
22: Add the cuts (28d) at \((x_t(\xi), y_t(\xi))\) for all \(\xi \in \Xi_k\).
23: Choose \(\Delta_{t+1} \in [\Delta_t, \bar{\Delta}]\).
24: end if
25: end loop

\((\mu_0(\xi), P_0(\xi))\) for all \(\xi \in \Xi_k\) and other parameters, including optimality gap \(\delta\) and serious step condition \(\eta\) (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 \(x_t(\xi)\) 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.

Theorem 5 (Theorem 2.4 in [17]) Assume that the subproblems (16) can be solved within a finite time. Algorithm 1 finds a sequence \(\{\mu_t(\xi), P_t(\xi)\}\) of dual iterates such that \(\lim_{t \to \infty} \sum_{\xi \in \Xi_k} \hat{D}_t = z(\Xi_k)\).

The convergence property is exactly the same as the BTR method developed in [17]. We remark that by Theorem 5, Algorithm 1 terminates after a finite number of iterations with \(\delta\)-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 [17]) and the bundle management strategy.

A branch-and-bound method can be incorporated to find a global optimal solution of DRMIP. For example, a branch-and-bound method has been proposed for branching on the nonanticipativity constraints in [9], which has been improved and implemented by [16]. In particular, the global optimal solutions are reported in [16] for all the SIPLIB test instances [3]. The same branching method, as in [9,16], 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 3, the Lagrangian dual problem (18) suffers from numerical instability with arbitrarily large bounds for $\alpha$ and $\beta_s$. Therefore, the numerical results from the problem (18) could not obtained or reported in this section.

6.1 Implementation

We have implemented Algorithm 1 in the open-source software package DSP [18]. Note that we have implemented only a serial version for our numerical experiments. In particular, we have used and modified the existing bundle TR method developed in [17] and implemented in DSP. The master problem (28) and scenario subproblems (16) were solved by CPLEX 12.8.0. 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., [27,24,18]) for its implicit regularization effect. 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 [8] with one additional file to specify reference scenarios $\hat{\xi}_s$, their probabilities $\hat{\rho}_s$, and the Wasserstein size $\epsilon$. 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 discretization $\Xi_k$, we set $\mu_0(\xi) = c/k$ for $\xi \in \Xi_k$, and $P_0(\xi) = \hat{\rho}_s$ if $\hat{\xi}_s = \xi$; otherwise, $P_0(\xi) = 0$ for $\xi \in \Xi_k$. Note that the initial dual value is feasible to the Lagrangian dual problem (15). We also set the parameters $\rho = 3$, $\bar{\rho} = 4$, $\Delta = 10^{-2}$, $\bar{\Delta} = 10^4$, $\Delta_0 = 0.1$, $\eta = 10^{-4}$, and $\delta = 10^{-5}$. We used $\ell_\infty$-norm for the TR constraints (28e) and (28f). At each iteration, we choose only one first-stage solution $x_t(\xi)$ 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 the support $\Xi$ of uncertain scenarios.

Let $x_{it}$ 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 $\xi$, indicated by variable $y_{ijt}(\xi) \in \{0, 1\}$. Let $a_{it}$ and $b_{it}$ respectively be the variable and fixed cost of resource $i$ at time $t$, and let $c_{ijt}(\xi)$ be the cost of assigning resource $i$ to task $j$ at time $t$ under scenario $\xi$. We also denote by $d_{jt}(\xi)$ the task $j$ requirement at time $t$ under scenario $\xi$. As a result, the problem is formulated as follows.

$$\min_{x,u,y(\xi)} \sum_{i=1}^{m} \sum_{t=1}^{T} (a_{it}x_{it} + b_{it}u_{it}) + \max_{P \in P} \mathbb{E}_P \left[ \sum_{i=0}^{m} \sum_{j=1}^{n} \sum_{t=1}^{T} c_{ijt}(\xi)y_{ijt}(\xi) \right]$$

s.t. 
- $x_{it} - u_{it} \leq 0 \quad \forall i,t,$
- $-\sum_{\tau=1}^{t} x_{i\tau} + \sum_{j=1}^{n} d_{jt}(\xi)y_{ijt}(\xi) \leq 0 \quad \forall i,t,\xi,$
- $\sum_{i=0}^{m} y_{ijt}(\xi) = 1 \quad \forall j,t,\xi,$
- $x_{it} \geq 0, \quad u_{it} \in \{0, 1\}, \quad y_{ijt}(\xi) \in \{0, 1\}, \quad \forall i,j,t,\xi.$

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 $|\xi|$ in the table the dimension of the uncertain parameters. The instances are named with the number $k$ of discretization points and the Wasserstein size $\epsilon$. 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 $s = 1, \ldots, N$. 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.

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 $c_{ijt}(\xi)$ and $d_{jt}(\xi)$ are defined by $[5, 10]$ and $[0.5, 1.5]$, respectively, for every $i > 0, j, t$. The support of $c_{0jt}(\xi)$ is defined by $[500, 1000]$ for every $j, t$. That is, the values can be anything randomly in the given ranges. For each $k \in \{20, 50, 100, 200, 300\}$, the discretization $\Xi_k$ of the support contains the reference scenarios $\hat{\xi}_s$, which ensures the feasibility of the Lagrangian dual prob-
Table 1: Description of test problem instances

<table>
<thead>
<tr>
<th>Name</th>
<th>m</th>
<th>n</th>
<th>T</th>
<th></th>
<th></th>
<th>Rows</th>
<th>Cols</th>
<th>Ints</th>
<th>Rows</th>
<th>Cols</th>
<th>Ints</th>
</tr>
</thead>
<tbody>
<tr>
<td>dcap233k,ε</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td>36</td>
<td>7</td>
<td>12</td>
<td>6</td>
<td>15</td>
<td>27</td>
</tr>
<tr>
<td>dcap243k,ε</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td>48</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>18</td>
<td>36</td>
</tr>
<tr>
<td>dcap332k,ε</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td>30</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>dcap342k,ε</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td>40</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>14</td>
<td>32</td>
</tr>
</tbody>
</table>

Moreover, we generated the support discretizations such that $\Xi_k \subset \Xi_{k'}$ for any $k < k'$. This allowed us to demonstrate the weak discretizability of the Lagrangian dual problem. 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.

We found the Lagrangian dual bounds (LB) and upper bounds (UB) for all 80 problem instances. Among the 80 problem instances, 76 instances result in an optimality gap less than 1%, of which 19 instances result in an optimality gap less than 0.01%. The results imply that the Lagrangian dual problem (15) is effective for finding the solutions of DRMIP with very small optimality gaps. The number of iterations ranges from 47 to 351. The total solution time ranges from 24 to 4948 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 [18].

Figure 1 shows the changes of total solution time as the Wasserstein distance limit $\epsilon$ increases for each $k \in \{20, 50, 100, 200, 300\}$. Significant increases in the total solution time are observed with the Wasserstein distance limit $\epsilon$ for the problem instances with larger numbers of scenarios. The reason is that the size of the DRMIP problem (7) increases with the number of discretization points and also results in more Lagrangian dual subproblems (16). In addition, since the Wasserstein ambiguity set (3) is larger with larger values of $\epsilon$ and $k$, the dual decomposition method tends to require more iterations before termination.

Figure 2 shows the Lagrangian dual bounds $z_{LD}(\Xi_k)$ of the test instances by increasing the Wasserstein distance limit $\epsilon$ for each $k \in \{20, 50, 100, 200, 300\}$. We recall that the Lagrangian dual bound may increase with larger Wasserstein ambiguity set by $\epsilon$ (i.e., being more robust to the uncertainty of our empirical distribution). However, the problem instances with smaller numbers of discretization points results in the lower bound being significantly lower than the instances with larger numbers of discretization points. This result implies that a sufficiently large number of discretization points are significant
Table 2: Numerical results from the dual decomposition solutions for dcap233_{k,\epsilon} and dcap243_{k,\epsilon} instances for \( k \in \{20, 50, 100, 200, 300\} \) and \( \epsilon \in \{1, 100, 500, 1000\} \).

<table>
<thead>
<tr>
<th>Name</th>
<th>Iteration</th>
<th>UB</th>
<th>LB</th>
<th>Gap (%)</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dcap233_{20,1}</td>
<td>52</td>
<td>2648.64</td>
<td>2648.57</td>
<td>0.00</td>
<td>29</td>
</tr>
<tr>
<td>dcap233_{20,100}</td>
<td>97</td>
<td>2720.53</td>
<td>2720.12</td>
<td>0.02</td>
<td>56</td>
</tr>
<tr>
<td>dcap233_{20,500}</td>
<td>161</td>
<td>2741.32</td>
<td>2741.04</td>
<td>0.01</td>
<td>81</td>
</tr>
<tr>
<td>dcap233_{20,1000}</td>
<td>151</td>
<td>2741.32</td>
<td>2741.05</td>
<td>0.01</td>
<td>79</td>
</tr>
<tr>
<td>dcap233_{50,1}</td>
<td>78</td>
<td>2648.64</td>
<td>2648.55</td>
<td>0.00</td>
<td>124</td>
</tr>
<tr>
<td>dcap233_{50,100}</td>
<td>106</td>
<td>2722.73</td>
<td>2720.16</td>
<td>0.09</td>
<td>168</td>
</tr>
<tr>
<td>dcap233_{50,500}</td>
<td>144</td>
<td>2782.49</td>
<td>2779.99</td>
<td>0.09</td>
<td>228</td>
</tr>
<tr>
<td>dcap233_{50,1000}</td>
<td>126</td>
<td>2822.79</td>
<td>2814.41</td>
<td>0.30</td>
<td>190</td>
</tr>
<tr>
<td>dcap233_{100,1}</td>
<td>120</td>
<td>2648.74</td>
<td>2648.66</td>
<td>0.00</td>
<td>357</td>
</tr>
<tr>
<td>dcap233_{100,100}</td>
<td>126</td>
<td>2775.34</td>
<td>2774.97</td>
<td>0.01</td>
<td>408</td>
</tr>
<tr>
<td>dcap233_{100,500}</td>
<td>146</td>
<td>3260.09</td>
<td>3258.58</td>
<td>0.05</td>
<td>446</td>
</tr>
<tr>
<td>dcap233_{100,1000}</td>
<td>132</td>
<td>3546.85</td>
<td>3526.07</td>
<td>0.59</td>
<td>349</td>
</tr>
<tr>
<td>dcap233_{200,1}</td>
<td>149</td>
<td>2648.91</td>
<td>2648.65</td>
<td>0.01</td>
<td>895</td>
</tr>
<tr>
<td>dcap233_{200,100}</td>
<td>152</td>
<td>2775.59</td>
<td>2774.96</td>
<td>0.02</td>
<td>1068</td>
</tr>
<tr>
<td>dcap233_{200,500}</td>
<td>181</td>
<td>3260.34</td>
<td>3258.65</td>
<td>0.05</td>
<td>1185</td>
</tr>
<tr>
<td>dcap233_{200,1000}</td>
<td>177</td>
<td>3577.54</td>
<td>3549.00</td>
<td>0.80</td>
<td>1049</td>
</tr>
<tr>
<td>dcap233_{300,1}</td>
<td>143</td>
<td>2648.88</td>
<td>2648.66</td>
<td>0.01</td>
<td>1485</td>
</tr>
<tr>
<td>dcap233_{300,100}</td>
<td>200</td>
<td>2775.57</td>
<td>2774.99</td>
<td>0.02</td>
<td>2006</td>
</tr>
<tr>
<td>dcap233_{300,500}</td>
<td>106</td>
<td>3260.32</td>
<td>3256.27</td>
<td>0.12</td>
<td>1194</td>
</tr>
<tr>
<td>dcap233_{300,1000}</td>
<td>228</td>
<td>3579.31</td>
<td>3549.01</td>
<td>0.85</td>
<td>2211</td>
</tr>
<tr>
<td>dcap243_{20,1}</td>
<td>47</td>
<td>3839.04</td>
<td>3838.78</td>
<td>0.04</td>
<td>44</td>
</tr>
<tr>
<td>dcap243_{20,100}</td>
<td>109</td>
<td>3981.40</td>
<td>3981.38</td>
<td>0.00</td>
<td>94</td>
</tr>
<tr>
<td>dcap243_{20,500}</td>
<td>174</td>
<td>4122.40</td>
<td>4119.15</td>
<td>0.08</td>
<td>140</td>
</tr>
<tr>
<td>dcap243_{20,1000}</td>
<td>120</td>
<td>4122.40</td>
<td>4119.17</td>
<td>0.08</td>
<td>90</td>
</tr>
<tr>
<td>dcap243_{50,1}</td>
<td>70</td>
<td>3838.79</td>
<td>3838.77</td>
<td>0.00</td>
<td>165</td>
</tr>
<tr>
<td>dcap243_{50,100}</td>
<td>70</td>
<td>3981.40</td>
<td>3981.38</td>
<td>0.00</td>
<td>229</td>
</tr>
<tr>
<td>dcap243_{50,500}</td>
<td>97</td>
<td>4122.47</td>
<td>4119.15</td>
<td>0.08</td>
<td>203</td>
</tr>
<tr>
<td>dcap243_{50,1000}</td>
<td>151</td>
<td>4122.40</td>
<td>4119.14</td>
<td>0.08</td>
<td>299</td>
</tr>
<tr>
<td>dcap243_{100,1}</td>
<td>95</td>
<td>3838.79</td>
<td>3838.78</td>
<td>0.00</td>
<td>227</td>
</tr>
<tr>
<td>dcap243_{100,100}</td>
<td>127</td>
<td>3981.40</td>
<td>3981.39</td>
<td>0.00</td>
<td>600</td>
</tr>
<tr>
<td>dcap243_{100,500}</td>
<td>136</td>
<td>4309.80</td>
<td>4308.86</td>
<td>0.02</td>
<td>619</td>
</tr>
<tr>
<td>dcap243_{100,1000}</td>
<td>124</td>
<td>4351.98</td>
<td>4326.84</td>
<td>0.58</td>
<td>480</td>
</tr>
<tr>
<td>dcap243_{200,1}</td>
<td>103</td>
<td>3838.79</td>
<td>3838.78</td>
<td>0.00</td>
<td>1001</td>
</tr>
<tr>
<td>dcap243_{200,100}</td>
<td>122</td>
<td>3987.58</td>
<td>3987.56</td>
<td>0.00</td>
<td>1190</td>
</tr>
<tr>
<td>dcap243_{200,500}</td>
<td>167</td>
<td>4413.76</td>
<td>4410.26</td>
<td>0.08</td>
<td>1594</td>
</tr>
<tr>
<td>dcap243_{200,1000}</td>
<td>208</td>
<td>4547.79</td>
<td>4517.66</td>
<td>0.66</td>
<td>1887</td>
</tr>
<tr>
<td>dcap243_{300,1}</td>
<td>120</td>
<td>3838.76</td>
<td>3838.73</td>
<td>0.00</td>
<td>1685</td>
</tr>
<tr>
<td>dcap243_{300,100}</td>
<td>127</td>
<td>3987.55</td>
<td>3987.54</td>
<td>0.00</td>
<td>2128</td>
</tr>
<tr>
<td>dcap243_{300,500}</td>
<td>159</td>
<td>4429.27</td>
<td>4427.96</td>
<td>0.03</td>
<td>2451</td>
</tr>
<tr>
<td>dcap243_{300,1000}</td>
<td>351</td>
<td>4760.59</td>
<td>4766.53</td>
<td>0.06</td>
<td>4948</td>
</tr>
</tbody>
</table>

To achieve statistically robust solution. We 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 \( \Omega \) without much increase in solution time. In fact, the strong scalability of the method has been shown for SMIP [18,17].
Table 3: Numerical results from the dual decomposition solutions for $\text{dcap}_{332,k,\epsilon}$ and $\text{dcap}_{342,k,\epsilon}$ instances for $k \in \{20, 50, 100, 200, 300\}$ and $\epsilon \in \{1, 100, 500, 1000\}$.

<table>
<thead>
<tr>
<th>Name</th>
<th>Iteration</th>
<th>UB</th>
<th>LB</th>
<th>Gap (%)</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{dcap}_{332,20_1}$</td>
<td>53</td>
<td>2106.38</td>
<td>2098.83</td>
<td>0.36</td>
<td>24</td>
</tr>
<tr>
<td>$\text{dcap}_{332,20_100}$</td>
<td>124</td>
<td>2182.82</td>
<td>2175.09</td>
<td>0.35</td>
<td>53</td>
</tr>
<tr>
<td>$\text{dcap}_{332,20_500}$</td>
<td>128</td>
<td>2210.87</td>
<td>2187.50</td>
<td>1.06</td>
<td>56</td>
</tr>
<tr>
<td>$\text{dcap}_{332,20_1000}$</td>
<td>117</td>
<td>2196.08</td>
<td>2187.50</td>
<td>0.39</td>
<td>52</td>
</tr>
<tr>
<td>$\text{dcap}_{332,50_1}$</td>
<td>85</td>
<td>2106.23</td>
<td>2099.30</td>
<td>0.33</td>
<td>100</td>
</tr>
<tr>
<td>$\text{dcap}_{332,50_100}$</td>
<td>108</td>
<td>2263.27</td>
<td>2255.75</td>
<td>0.33</td>
<td>129</td>
</tr>
<tr>
<td>$\text{dcap}_{332,50_500}$</td>
<td>146</td>
<td>2853.40</td>
<td>2826.72</td>
<td>0.94</td>
<td>157</td>
</tr>
<tr>
<td>$\text{dcap}_{332,50_1000}$</td>
<td>151</td>
<td>2905.09</td>
<td>2847.77</td>
<td>1.97</td>
<td>166</td>
</tr>
<tr>
<td>$\text{dcap}_{332,100_1}$</td>
<td>87</td>
<td>2106.32</td>
<td>2099.39</td>
<td>0.33</td>
<td>223</td>
</tr>
<tr>
<td>$\text{dcap}_{332,100_100}$</td>
<td>131</td>
<td>2274.55</td>
<td>2266.19</td>
<td>0.37</td>
<td>339</td>
</tr>
<tr>
<td>$\text{dcap}_{332,100_500}$</td>
<td>122</td>
<td>2909.30</td>
<td>2887.54</td>
<td>0.39</td>
<td>308</td>
</tr>
<tr>
<td>$\text{dcap}_{332,100_1000}$</td>
<td>160</td>
<td>3017.40</td>
<td>2951.71</td>
<td>2.18</td>
<td>375</td>
</tr>
<tr>
<td>$\text{dcap}_{332,200_1}$</td>
<td>88</td>
<td>2106.32</td>
<td>2099.39</td>
<td>0.33</td>
<td>509</td>
</tr>
<tr>
<td>$\text{dcap}_{332,200_100}$</td>
<td>157</td>
<td>2273.08</td>
<td>2266.19</td>
<td>0.30</td>
<td>900</td>
</tr>
<tr>
<td>$\text{dcap}_{332,200_500}$</td>
<td>170</td>
<td>2908.81</td>
<td>2885.87</td>
<td>0.44</td>
<td>1037</td>
</tr>
<tr>
<td>$\text{dcap}_{332,200_1000}$</td>
<td>286</td>
<td>3159.87</td>
<td>3152.62</td>
<td>0.23</td>
<td>1447</td>
</tr>
<tr>
<td>$\text{dcap}_{332,300_1}$</td>
<td>109</td>
<td>2107.59</td>
<td>2099.38</td>
<td>0.39</td>
<td>944</td>
</tr>
<tr>
<td>$\text{dcap}_{332,300_100}$</td>
<td>158</td>
<td>2274.14</td>
<td>2266.19</td>
<td>0.35</td>
<td>1543</td>
</tr>
<tr>
<td>$\text{dcap}_{332,300_500}$</td>
<td>150</td>
<td>2910.16</td>
<td>2895.99</td>
<td>0.49</td>
<td>1509</td>
</tr>
<tr>
<td>$\text{dcap}_{332,300_1000}$</td>
<td>280</td>
<td>3169.40</td>
<td>3152.99</td>
<td>0.52</td>
<td>2445</td>
</tr>
<tr>
<td>$\text{dcap}_{342,20_1}$</td>
<td>78</td>
<td>3173.29</td>
<td>3168.02</td>
<td>0.17</td>
<td>29</td>
</tr>
<tr>
<td>$\text{dcap}_{342,20_100}$</td>
<td>101</td>
<td>3288.29</td>
<td>3280.70</td>
<td>0.26</td>
<td>40</td>
</tr>
<tr>
<td>$\text{dcap}_{342,20_500}$</td>
<td>106</td>
<td>3406.87</td>
<td>3401.75</td>
<td>0.15</td>
<td>55</td>
</tr>
<tr>
<td>$\text{dcap}_{342,20_1000}$</td>
<td>149</td>
<td>3418.43</td>
<td>3418.43</td>
<td>0.00</td>
<td>75</td>
</tr>
<tr>
<td>$\text{dcap}_{342,50_1}$</td>
<td>53</td>
<td>3173.29</td>
<td>3168.02</td>
<td>0.17</td>
<td>81</td>
</tr>
<tr>
<td>$\text{dcap}_{342,50_100}$</td>
<td>101</td>
<td>3288.29</td>
<td>3280.70</td>
<td>0.23</td>
<td>149</td>
</tr>
<tr>
<td>$\text{dcap}_{342,50_500}$</td>
<td>128</td>
<td>3401.78</td>
<td>3401.72</td>
<td>0.00</td>
<td>185</td>
</tr>
<tr>
<td>$\text{dcap}_{342,50_1000}$</td>
<td>130</td>
<td>3418.43</td>
<td>3418.41</td>
<td>0.00</td>
<td>192</td>
</tr>
<tr>
<td>$\text{dcap}_{342,100_1}$</td>
<td>81</td>
<td>3173.18</td>
<td>3168.01</td>
<td>0.16</td>
<td>251</td>
</tr>
<tr>
<td>$\text{dcap}_{342,100_100}$</td>
<td>126</td>
<td>3284.37</td>
<td>3280.71</td>
<td>0.11</td>
<td>420</td>
</tr>
<tr>
<td>$\text{dcap}_{342,100_500}$</td>
<td>123</td>
<td>3488.03</td>
<td>3482.62</td>
<td>0.16</td>
<td>382</td>
</tr>
<tr>
<td>$\text{dcap}_{342,100_1000}$</td>
<td>225</td>
<td>3557.35</td>
<td>3549.94</td>
<td>0.21</td>
<td>682</td>
</tr>
<tr>
<td>$\text{dcap}_{342,200_1}$</td>
<td>120</td>
<td>3173.29</td>
<td>3168.01</td>
<td>0.17</td>
<td>799</td>
</tr>
<tr>
<td>$\text{dcap}_{342,200_100}$</td>
<td>168</td>
<td>3333.59</td>
<td>3324.28</td>
<td>0.31</td>
<td>1171</td>
</tr>
<tr>
<td>$\text{dcap}_{342,200_500}$</td>
<td>145</td>
<td>3919.34</td>
<td>3909.90</td>
<td>0.24</td>
<td>976</td>
</tr>
<tr>
<td>$\text{dcap}_{342,200_1000}$</td>
<td>224</td>
<td>4150.47</td>
<td>4144.97</td>
<td>0.13</td>
<td>1454</td>
</tr>
<tr>
<td>$\text{dcap}_{342,300_1}$</td>
<td>171</td>
<td>3173.27</td>
<td>3168.00</td>
<td>0.17</td>
<td>1767</td>
</tr>
<tr>
<td>$\text{dcap}_{342,300_100}$</td>
<td>169</td>
<td>3333.39</td>
<td>3324.28</td>
<td>0.27</td>
<td>1928</td>
</tr>
<tr>
<td>$\text{dcap}_{342,300_500}$</td>
<td>202</td>
<td>3916.18</td>
<td>3909.90</td>
<td>0.16</td>
<td>2376</td>
</tr>
<tr>
<td>$\text{dcap}_{342,300_1000}$</td>
<td>235</td>
<td>4200.30</td>
<td>4144.99</td>
<td>1.32</td>
<td>2434</td>
</tr>
</tbody>
</table>

### 6.4 Impact of weak discretization

In this section we discuss the behavior and impact of the discretization $\Xi_k$ of the support $\Xi$. Recall that $\Xi_k \subset \Xi_k'$ and thus that $z_{LD}(\Xi_k) \leq z_{LD}(\Xi_k')$ for $k < k'$ and $k, k' \in \{20, 50, 100, 200, 300\}$.

Figure 3 shows the nondecreasing values of the Lagrangian bound as the number of sample points increases. Moreover, for all the instances except for
The numerical results are consistent with the mathematical properties shown in Theorem 4 and Corollary 2. The plots of the dual bounds are flatter with smaller Wasserstein limit (e.g., $\epsilon \leq 100$). The implication is that a smaller number of discretization points may be sufficient if we are more confident with the empirical distribution (i.e., smaller Wasserstein set). Consequently, the numerical results show that the Lagrangian dual problem (15) is weakly discretizable. The discretization with a large number of points can effectively solve the original DRMIP with continuous support.

7 Summary and Directions of Future Work

We have developed the dual decomposition method for solving DRMIP with a discretization of the support. To this end, we first derived the deterministic equivalent form of DRMIP as an infinite-dimensional 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 $(\alpha, \beta_s)$, 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 \((\alpha, \beta_s)\). Note that exact column bounds are not given or cannot be obtained in general.

As of SMIP form, DRMIP adapts the existing properties and algorithms for SMIP, including the convergence properties of the SAA. In addition, we analyzed the structural properties of DRMIP and the Lagrangian dual with the view of infinite-dimensional programming. In particular, we proved that DRMIP is finitely reducible and thus weakly discretizable. We also proved that the Lagrangian dual of DRMIP is weakly discretizable. These properties imply that a sufficiently large number of discretization points can offer good approximation (or even exact) solutions to DRMIP. We used that to justify using the dual decomposition method for a discretized DRMIP, particularly with the strong scalability of the method, as shown in [18, 17].

In this paper, we have implemented a serial computation of the dual decomposition method for DRMIP. The existing dual decomposition framework in DSP was not readily available because of the different communication pattern required for the upper bounding problem (33), which requires a collective operation. The parallelization of the method can be achieved as future work. On the other hand, a dynamic discretization scheme would also be of interest as future work. As briefly described at the end of Section 4, one can develop

---

Fig. 2: Lagrangian dual bounds \(z_{LD}(\Xi_k)\) for the test instances under different values of the Wasserstein distance limit \(\epsilon\) for \(k \in \{20, 50, 100, 200, 300\}\).
Fig. 3: Lagrangian dual bounds $z_{LD}(\Xi_k)$ for the test instances under different values of the Wasserstein distance limit $\epsilon$ for $k \in \{20, 50, 100, 200, 300\}$.

a discretization method that systematically finds a sequence of superset discretization.

References

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).