

# A Novel Solution Methodology for Wasserstein-based Data-Driven Distributionally Robust Problems

Carlos Andrés Gamboa<sup>1</sup>, Davi Michel Valladão<sup>2</sup>, Alexandre Street<sup>3</sup>, and Tito Homem-de-Mello<sup>4</sup>

<sup>1</sup>Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil, caagamboaro@aluno.puc-rio.br

<sup>2</sup>Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil, davimv@puc-rio.br

<sup>3</sup>Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil, street@ele.puc-rio.br

<sup>4</sup>School of Business, Universidad Adolfo Ibáñez, Chile, tito.hmello@uai.cl

## Abstract

Distributionally robust optimization (DRO) is a mathematical framework to incorporate ambiguity over the actual data-generating probability distribution. Data-driven DRO problems based on the Wasserstein distance are of particular interest for their sound mathematical properties. For right-hand-sided uncertainty, however, existing methods rely on dual vertex enumeration rendering the problem intractable in practical applications. In this context, we study decomposition methods for two-stage data-driven Wasserstein-based DROs with right-hand-sided uncertainty and rectangular support. We propose a novel finite reformulation that explores the rectangular uncertainty support to develop and test three new different decomposition schemes: Column-Constraint Generation, Single-cut Benders and Multi-cut Benders. We compare the efficiency of the proposed methods with existing benchmarks for a unit commitment problem with 5, 14, and 54 thermal generators over a 24-hour uncertainty dimension. The numerical results show the superiority of the Column-Constraint Generation algorithm for this problem, in contrast to the Benders-like alternatives and existing methods.

**Index terms**— two-stage stochastic programming, decomposition methods, distributionally robust optimization

## 1 Introduction

Distributionally Robust Optimization (DRO) is a mathematical framework to incorporate ambiguity in the characterization of the true data-generating distribution. In general, this approach considers an ambiguity set  $\mathcal{P}$  to evaluate the worst-case for the expectation  $\mathbb{E}^{\tilde{P}}[Q(\mathbf{x}, \xi)]$  over all probability distributions<sup>1</sup>  $\tilde{P} \in \mathcal{P}$ . Then, for a DRO minimization problem, the optimal decision  $\mathbf{x} \in X$  is the one that minimizes the highest expected cost  $\mathbb{E}^{\tilde{P}}[Q(\mathbf{x}, \xi)]$  over all  $\tilde{P} \in \mathcal{P}$ .

Research on DRO models—formulations, algorithms and applications—has grown enormously in the past few years; a recent survey can be found in Rahimian and Mehrotra (2019). One particular setting that has received considerable attention in the literature is when the ambiguity set  $\mathcal{P}$  is defined as the set of distributions that are not “too far” from some reference distribution. Of course, such a notion requires defining an appropriate way to measure the distance between distributions. While there are multiple ways to measure such distance, the *Wasserstein* distance has been popular due to

its theoretical properties and practical performance. The thrust of the two-stage *data-driven* DRO with Wasserstein metric (DD-DRO-W) framework is to center the ambiguity set on the empirical distribution corresponding to the data and to set a radius around it to enclose the true data-generating distribution with a high confidence level (Mohajerin Esfahani and Kuhn 2018). The data-driven Wasserstein distance has been widely used in DRO partly because the associate ambiguity set collects both discrete and continuous probability distributions, even though the Wasserstein ball is centered around the empirical discrete distribution.

Despite their practical appeal, DD-DRO-W problems are hard to solve, but specific characteristics can be explored for computational tractability. For instance, by considering a compact support set for the uncertainty, Wozabal (2012) presents a finite-dimensional non-convex reformulation of the worst-case expectation problem, which has high computational burden and with no global optimality guarantee. For a continuous data-generating distribution supported on a polyhedron, Zhao (2014) presents a semi-infinite linear reformulation of a two-stage distributionally robust unit commitment problem for the integration of renewable energy over the data-driven Wasserstein ball. An exact decomposition scheme is proposed by Bansal et al. (2018), but considering an ambiguity set with only discrete probability distributions. In Hanasusanto and Kuhn (2018), the authors present a general conic programming reformulation for a two-stage distributionally robust optimization problem with Wasserstein-based ambiguity set. Particularly, for linear programming problems with right-hand-sided uncertainty, the authors present a tractable reformulation for the case of unbounded uncertainty supports, an unsuited assumption for a large range of applications. By assuming a compact support set for the uncertainty, Mohajerin Esfahani and Kuhn (2018) present tractable reformulations for a number of cases, except for linear programming problems with right-hand-sided uncertainty. For this case, the resulting reformulation requires pre-computing all dual vertices of the recourse problem, but such number grows exponentially with the size of the problem.

In this context, we study decomposition methods applied to two-stage DD-DRO-W with right-hand-sided uncertainty and *rectangular* support. As an alternative to the proposed dual vertex enumeration in Mohajerin Esfahani and Kuhn (2018), in Section 3, we propose a novel finite reformulation that explores the rectangular uncertainty support. We develop an exact decomposition (oracle-master) scheme based on the Column-and-Constraint Generation (C&CG) method. Moreover, by considering variations of the proposed scheme, we derive two other alternative decomposition methods, namely, Multi-cut Benders and Single-cut Benders. Whereas Benders’ methods consider a local linear approximation of the recourse function for a given scenario, the C&CG method tends to converge faster since co-optimizes first-stage and the recourse for the uncertainty realizations selected by the oracle problem.

It is important to mention that just prior to the submission of this paper we came across a paper by Duque et al. (2020)—which was developed independently of our work and very recently made available online—which solves a similar but different class of problems. Unlike our framework that considers all problems with uncertainty (with rectangular support) on the right-hand side, the authors do not allow randomness in the technology matrix, i.e., the matrix multiplying the first stage variable. On the other hand, Duque et al. (2020) consider problems in which the uncertainty has either bounded or unbounded support. They provide a master-oracle scheme that resembles our Benders Multi-cut method with a similar master but with a different MILP oracle reformulation, where the uncertain parameter does not appear as a coefficient of the first-stage variable. Also, Duque et al. (2020) do not propose an algorithm similar to C&CG which turns out to be the most efficient in our numerical experiments.

We compare the proposed C&CG method with two variants of the Benders algorithm applied to the newly proposed finite reformulation. Furthermore, we also compare the proposed methods with the existing formulation provided in Mohajerin Esfahani and Kuhn (2018) using the vertex enumeration algorithm available in Fukuda and Prodon (1995), Motzkin et al. (1953) and the decomposition method presented in Bansal et al. (2018). To do that, we present results for the unit commitment problem with 5, 14, and 54 thermal generators over a 24-hour uncertainty dimension.

We summarize below our main contributions to data-driven Wasserstein-based DRO problems with

rectangular-support right-hand-side uncertainty:

- We propose a novel finite reformulation that explores the rectangular uncertainty support.
- We develop a novel master-oracle decomposition framework based on a new and exact MILP-based oracle subproblem.
- We develop three decomposition methods, namely, Column-Constraint Generation, Single-cut Benders, and Multi-cut Benders.
- We illustrate the computational performance of the proposed methods for a unit commitment problem with 5, 14, and 54 thermal generators over a 24-hour uncertainty dimension.

The remainder of this paper is organized as follows: In Section 2, we introduce the definition of the Wasserstein ball and present the two-stage distributionally robust optimization problem with right-hand-side uncertainty under a data-driven Wasserstein ambiguity set. In Section 3, we develop a new tractable reformulation of the DD-DRO-W problem suitable to be solved by decomposition methods. The solution methodology to assess the extensive form of the derived tractable reformulation is presented in Section 4. In Section 5, we present numerical experiments for assessing the proposed solution methodology and we analyze the computational performance of the develop algorithmic schemes. Finally, we present concluding remarks in Section 6.

## 2 Problem statement

In this section, we introduce the data-driven Wasserstein-based ambiguity set (Wasserstein ball). Next, we present the data-driven two-stage distributionally robust optimization problem with right-hand-side uncertainty and a Wasserstein-based ambiguity set.

### 2.1 Wasserstein-based ambiguity set

We assume that we can access a finite data-set of training samples  $\{\widehat{\xi}_n\}_{n \leq N}$  that are generated by the true (but unknown) probability distribution. We consider the reference distribution

$$\widehat{P}_N = \frac{1}{N} \sum_{n=1}^N \delta_{\widehat{\xi}_n}, \quad (1)$$

where  $\frac{1}{N}$  is the probability mass of the point  $\widehat{\xi}_n$  and  $\delta_{\widehat{\xi}_n}$  denotes the Dirac's delta function that concentrates unit mass at  $\widehat{\xi}_n$ .

For a given set  $\Xi \subseteq \mathbb{R}^{d_\xi}$ , let  $\mathcal{M}(\Xi)$  be the space of probability distributions  $\widetilde{P}$  supported on  $\Xi$  such that  $\mathbb{E}^{\widetilde{P}}[\|\xi\|] < \infty$ . For a given  $\delta > 0$ , the data-driven Wasserstein ball is defined by

$$\mathbb{B}_\delta(\widehat{P}_N) = \{\widetilde{P} \in \mathcal{M}(\Xi) : d_w(\widetilde{P}, \widehat{P}_N) \leq \delta\} \quad (2)$$

where  $d_w(\widetilde{P}, \widehat{P}_N)$  denotes the Wasserstein distance between the probability distribution  $\widetilde{P} \in \mathcal{M}(\Xi)$  and the reference distribution  $\widehat{P}_N$ , which is defined as follows (Kantorovich and Rubinshtein 1958):

$$d_w(\widetilde{P}, \widehat{P}_N) := \inf \left\{ \int_{\Xi^2} \|\xi - \widehat{\xi}_n\| d\Pi(\xi, \widehat{\xi}_n) : \begin{array}{l} \Pi \text{ is a joint distribution of } \xi \text{ and } \widehat{\xi}_n \\ \text{with marginals } \widetilde{P} \text{ and } \widehat{P}_N, \text{ respectively} \end{array} \right\} \quad (3)$$

where  $\|\cdot\|$  represents an arbitrary norm on  $\mathbb{R}^{d_\xi}$ . It is well known that the decision variable  $\Pi$  in (3) can be viewed as a transportation plan for moving a probability mass described by distribution  $\widetilde{P}$  to another one described by  $\widehat{P}_N$ ; see, e.g., Rachev and Rüschendorf (1998).

In general, computing the Wasserstein distance when both probability distributions are not finite is #P-hard (see Kuhn et al. (2019)); despite that shortcoming, this metric is often used in Distributionally Robust Optimization (DRO) because of its nice properties—in particular, the ambiguity set defined by the Wasserstein distance (Wasserstein ball) contains all probability distributions (continuous and discrete) whose Wasserstein distance to the reference distribution  $\widehat{P}_N$  is less or equal to  $\delta$ .

By using the Wasserstein ball (2), the two-stage distributionally robust optimization problem with right-hand-side uncertainty can be defined as follows:

$$\min_{\mathbf{x} \in X} \mathbf{c}^\top \mathbf{x} + \sup_{\widetilde{P} \in \mathbb{B}_\delta(\widehat{P}_N)} \mathbb{E}^{\widetilde{P}}[Q(\mathbf{x}, \xi)], \quad (4)$$

where  $X \subseteq \mathbb{R}^{d_x}$  is the set of feasible first-stage decisions  $\mathbf{x}$ ,  $\mathbf{c} \in \mathbb{R}^{d_x}$  and  $Q : X \times \Xi \rightarrow \mathbb{R}$  is the recourse function (second-stage problem) that depends on the first-stage decision  $\mathbf{x}$  and the uncertain variable  $\xi$ . As usual, we assume a recourse function  $Q(\mathbf{x}, \xi)$  given by the optimal value of a parametric linear optimization problem, i.e.,<sup>2</sup>

$$Q(\mathbf{x}, \xi) = \min_{\mathbf{y} \geq 0} \mathbf{q}^\top \mathbf{y} \quad (5a)$$

$$\text{s.t. } \mathbf{W}\mathbf{y} = \mathbf{H}(\mathbf{x})\xi + \mathbf{r}(\mathbf{x}). \quad (5b)$$

In this problem,  $\mathbf{H}(\mathbf{x}) \in \mathbb{R}^{m_y \times d_\xi}$  and  $\mathbf{r}(\mathbf{x}) \in \mathbb{R}^{m_y}$  represent a decision-dependent matrix and vector, respectively. We assume that  $\mathbf{H} : X \rightarrow \mathbb{R}^{m_y \times d_\xi}$  and  $\mathbf{r} : X \rightarrow \mathbb{R}^{m_y}$  are affine functions of  $\mathbf{x}$ . Finally,  $\mathbf{q} \in \mathbb{R}^{d_y}$  and  $\mathbf{W} \in \mathbb{R}^{m_y \times d_y}$ . We call henceforth the inner problem  $\sup_{\widetilde{P} \in \mathbb{B}_\delta(\widehat{P}_N)} \mathbb{E}^{\widetilde{P}}[Q(\mathbf{x}, \xi)]$  as the worst-case expectation problem.

As shown in Mohajerin Esfahani and Kuhn (2018) and Zhao and Guan (2018), if (4) is such that  $\sup_{\xi \in \Xi} |Q(\mathbf{x}, \xi)| < \infty$  for all  $\mathbf{x} \in X$ , i.e., has relatively complete recourse with bounded recourse value, then it is equivalent to the semi-infinite optimization problem

$$\min_{\mathbf{x}, \lambda, s_n} \mathbf{c}^\top \mathbf{x} + \lambda \delta + \frac{1}{N} \sum_{n=1}^N s_n \quad (6a)$$

$$\text{s.t. } Q(\mathbf{x}, \xi) - \lambda \|\xi - \widehat{\xi}_n\| \leq s_n, \quad \forall n \leq N, \forall \xi \in \Xi \quad (6b)$$

$$\lambda \geq 0, \quad (6c)$$

$$\mathbf{x} \in X. \quad (6d)$$

In the next section, by considering the primal  $\ell_1$ -norm in (6b), we propose a (finite) linear programming reformulation for the semi-infinite problem (6) exploring the particular case of rectangular uncertainty support. We argue that this development is also valid for the  $\ell_\infty$ -norm.

### 3 A new finite reformulation for a rectangular uncertainty support

In this section, we derive a finite reformulation for the semi-infinite problem (6) by considering the  $\ell_1$ -norm (or equivalently the  $\ell_\infty$ -norm). We start by stating formally our assumptions:

**Assumption 1.** We assume that: (i) the second stage-problem belongs to the class of parametric linear programs with right-hand-side uncertainty as in (5); (ii) the set  $\{\mathbf{y} \in \mathbb{R}_+^{d_y} : \mathbf{W}\mathbf{y} = \mathbf{H}(\mathbf{x})\xi + \mathbf{r}(\mathbf{x})\}$  is nonempty and the optimal value of (5) is bounded for all  $\mathbf{x} \in X$  and  $\xi \in \Xi$ , and (iii) the uncertainty support  $\Xi \subseteq \mathbb{R}^{d_\xi}$  is a hypercube, i.e.,  $\Xi = \times_{i=1}^{d_\xi} [a_i, b_i]$ .

Within this framework, for any given  $\mathbf{x} \in X$ , the constraint (6b) is equivalent to

$$\sup_{\xi \in \Xi} \left( Q(\mathbf{x}, \xi) - \lambda \|\xi - \widehat{\xi}_n\|_1 \right) \leq s_n, \quad \forall n \leq N. \quad (7)$$

By representing the norm with linear inequalities and considering the dual formulation of  $Q(\mathbf{x}, \xi)$ , the left-hand side of constraint (7) is equivalent to the following nonlinear optimization problem (with bilinear term  $\theta^\top \xi$ ):

$$\begin{aligned}
& \sup_{\theta, \alpha, \xi} \quad \theta^\top \mathbf{r}(\mathbf{x}) + [\mathbf{H}^\top(\mathbf{x})\theta]^\top \xi - \lambda \mathbf{1}^\top \alpha \\
& \text{s.t.} \quad \alpha \geq \xi - \widehat{\xi}_n, \\
& \quad \alpha \geq \widehat{\xi}_n - \xi, \\
& \quad \mathbf{W}^\top \theta \leq \mathbf{q}, \\
& \quad \xi \in \Xi.
\end{aligned} \tag{8}$$

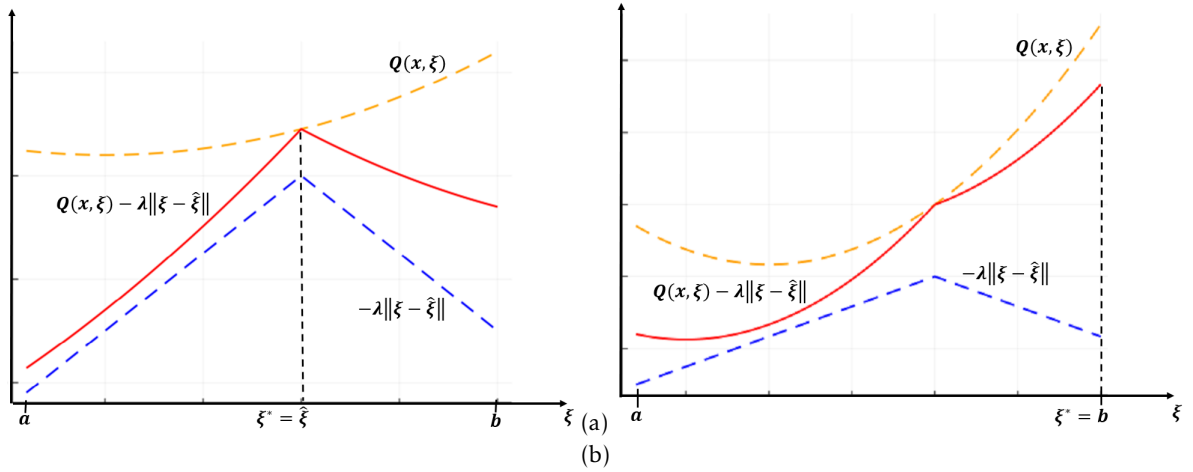


Figure 1: Illustrative example of the optimal solution  $\xi^*$  of (8) equals to  $\widehat{\xi}_n$  (a) and extreme point of the uncertainty support (b).

As illustrated by Fig. 1 and stated in Proposition 2, under Assumption 1, for a given scenario  $\widehat{\xi}_n$ ,  $n \leq N$ , the  $i$ -th component of the optimal solution  $\xi^*$  of the equivalent inner problem (8) is an extreme point of the interval  $[a_i, b_i]$  or the  $i$ -th component of the nominal value  $(\widehat{\xi}_n)_i$ .

**Proposition 2.** *Suppose Assumption 1 holds. Then, there exists an optimal solution  $\xi^*$  to (8) such that  $\xi_i^* \in \{(\widehat{\xi}_n)_i, a_i, b_i\}$ ,  $i = 1, \dots, d_\xi$ .*

*Proof.* Proof: Consider the rightmost problem in (8), and write it as

$$\sup_{\theta: \mathbf{W}^\top \theta \leq \mathbf{q}} \theta^\top \mathbf{r}(\mathbf{x}) + J(\theta)$$

where

$$J(\theta) := \sup_{\alpha \geq 0} -\lambda \mathbf{1}^\top \alpha + V(\theta, \alpha), \tag{9}$$

and  $V(\theta, \alpha)$  is given by

$$\begin{aligned}
& \max_{\xi \in \Xi} \quad [\mathbf{H}^\top(\mathbf{x})\theta]^\top \xi \\
& \text{s.t.} \quad \xi \leq \widehat{\xi}_n + \alpha, \\
& \quad \xi \geq \widehat{\xi}_n - \alpha,
\end{aligned} \tag{10}$$

i.e.,

$$V(\theta, \alpha) = \max_{\xi \in \Xi} \mathbf{w}(\theta)^\top \xi - \mathbb{I}_C(\xi, \alpha) \tag{11}$$

where  $\mathbf{w}(\boldsymbol{\theta}) := \mathbf{H}^\top(\mathbf{x})\boldsymbol{\theta}$ , and  $\mathbb{I}_C(\boldsymbol{\xi}, \boldsymbol{\alpha})$  is the indicator function of the set

$$C := \{(\boldsymbol{\xi}, \boldsymbol{\alpha}) : |\xi_i - \widehat{(\boldsymbol{\xi}_n)_i}| \leq \alpha_i, \quad i = 1, \dots, d_\xi\}.$$

Since the set  $C$  is convex, it follows that  $-\mathbb{I}_C(\boldsymbol{\xi}, \boldsymbol{\alpha})$  is concave and therefore  $V(\boldsymbol{\theta}, \boldsymbol{\alpha})$  is concave in  $\boldsymbol{\alpha}$ . Moreover, it is easy to see that the problem in (11) can be decomposed per coordinate  $\xi_i$ , i.e.,  $V(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=1}^{d_\xi} V_i(\boldsymbol{\theta}, \alpha_i)$ , where

$$V_i(\boldsymbol{\theta}, \alpha_i) = \max\{w_i(\boldsymbol{\theta})\xi_i : |\xi_i - \widehat{(\boldsymbol{\xi}_n)_i}| \leq \alpha_i, \xi_i \in [a_i, b_i]\}. \quad (12)$$

The solution to (12) is trivial to determine: if  $w_i(\boldsymbol{\theta}) > 0$ , then  $\xi_i^*(\alpha_i) = \min\{b_i, \widehat{(\boldsymbol{\xi}_n)_i} + \alpha_i\}$ ; otherwise,  $\xi_i^*(\alpha_i) = \max\{a_i, \widehat{(\boldsymbol{\xi}_n)_i} - \alpha_i\}$ . It follows that the function  $J(\boldsymbol{\theta})$  defined in (9) can be decomposed as

$$J(\boldsymbol{\theta}) := \sum_{i=1}^{d_\xi} \sup_{\alpha_i \geq 0} \begin{cases} -\lambda\alpha_i + w_i(\boldsymbol{\theta}) \min\{b_i, \widehat{(\boldsymbol{\xi}_n)_i} + \alpha_i\} & \text{if } w_i(\boldsymbol{\theta}) > 0 \\ -\lambda\alpha_i + w_i(\boldsymbol{\theta}) \max\{a_i, \widehat{(\boldsymbol{\xi}_n)_i} - \alpha_i\} & \text{otherwise.} \end{cases} \quad (13)$$

We see that, when  $w_i(\boldsymbol{\theta}) > 0$ , the expression inside the sup in (13) is maximized at  $\alpha_i = 0$  if  $\lambda > w_i(\boldsymbol{\theta})$ , and it is maximized at all values of  $\alpha_i \geq b_i - \widehat{(\boldsymbol{\xi}_n)_i}$  if  $\lambda \leq w_i(\boldsymbol{\theta})$ . Similarly, when  $w_i(\boldsymbol{\theta}) \leq 0$ , the expression inside the sup in (13) is maximized at  $\alpha_i = 0$  if  $\lambda > |w_i(\boldsymbol{\theta})|$ , and it is maximized at all values of  $\alpha_i \geq \widehat{(\boldsymbol{\xi}_n)_i} - a_i$  if  $\lambda \leq |w_i(\boldsymbol{\theta})|$ . The value of  $J(\boldsymbol{\theta})$  is then given by

$$J(\boldsymbol{\theta}) = \sum_{i=1}^{d_\xi} \begin{cases} -w_i(\boldsymbol{\theta})\widehat{(\boldsymbol{\xi}_n)_i} & \text{if } \lambda > |w_i(\boldsymbol{\theta})| \\ -\lambda(b_i - \widehat{(\boldsymbol{\xi}_n)_i}) + w_i(\boldsymbol{\theta})b_i & \text{if } \lambda \leq |w_i(\boldsymbol{\theta})| \text{ and } w_i(\boldsymbol{\theta}) > 0 \\ -\lambda(\widehat{(\boldsymbol{\xi}_n)_i} - a_i) + w_i(\boldsymbol{\theta})a_i & \text{if } \lambda \leq |w_i(\boldsymbol{\theta})| \text{ and } w_i(\boldsymbol{\theta}) \leq 0 \end{cases}$$

Moreover, by substituting the optimal values of  $\alpha_i$  found above into (12), we conclude that an optimal solution to the maximization problem in (12) is given by

$$\begin{cases} \xi_i^* = \widehat{(\boldsymbol{\xi}_n)_i} & \text{if } \lambda > |w_i(\boldsymbol{\theta})| \\ \xi_i^* = b_i & \text{if } \lambda \leq |w_i(\boldsymbol{\theta})| \text{ and } w_i(\boldsymbol{\theta}) > 0 \\ \xi_i^* = a_i & \text{if } \lambda \leq |w_i(\boldsymbol{\theta})| \text{ and } w_i(\boldsymbol{\theta}) \leq 0. \end{cases}$$

That is,  $\xi_i^* \in \{\widehat{(\boldsymbol{\xi}_n)_i}, a_i, b_i\}$  regardless of the value of  $\boldsymbol{\theta}$ . We conclude that there always exists an optimal solution  $\boldsymbol{\xi}^*$  to (8) such that  $\xi_i^* \in \{\widehat{(\boldsymbol{\xi}_n)_i}, a_i, b_i\}$ ,  $i = 1, \dots, d_\xi$ .  $\square$

Let us introduce the notation  $\widehat{\Xi}_n = \times_{i=1}^{d_\xi} \{a_i, \widehat{(\boldsymbol{\xi}_n)_i}, b_i\} = \{\boldsymbol{\xi}_\ell^*\}_{\ell \in \mathcal{L}_n}$  and  $\mathcal{L}_n = \{1, \dots, |\widehat{\Xi}_n|\}$ , for all  $n = 1, \dots, N$ . According to **Proposition 2**, the set  $\widehat{\Xi}_n$  comprises the (eligible) candidates for optimal solution of the sup problem on the left side of (7) for any feasible pair  $(\mathbf{x}, \lambda)$ , and for each scenario  $\widehat{\boldsymbol{\xi}}_n$ ,  $n = 1, \dots, N$ . With this notation at hand, by replacing the infinite set  $\Xi$  with the finite set  $\widehat{\Xi}_n$ , for all  $n = 1, \dots, N$ , in the constraint (6b), the problem (6) reduces to the following linear program:

$$\min_{\mathbf{x}, \lambda, s_n} \mathbf{c}^\top \mathbf{x} + \lambda\delta + \frac{1}{N} \sum_{n=1}^N s_n \quad (14a)$$

$$\text{s.t. } Q(\mathbf{x}, \boldsymbol{\xi}_\ell^*) - \lambda \|\boldsymbol{\xi}_\ell^* - \widehat{\boldsymbol{\xi}}_n\|_1 \leq s_n, \quad \forall \ell \in \mathcal{L}_n, \forall n \leq N, \quad (14b)$$

$$\lambda \geq 0, \quad (14c)$$

$$\mathbf{x} \in \mathbf{X}, \quad (14d)$$

which scales with the number of candidate solutions in  $\mathcal{L}_n$ , for all  $n = 1, \dots, N$ . In the next section, we propose decomposition schemes to handle the extensive linear program (14).

Note that **Proposition 2** also establishes an equivalence of the worst-case expectation problem in the left-hand side of constraint (7) with the distribution separation problem (see Bansal et al. (2018)) based

on the Wasserstein set  $\widehat{\mathbb{B}}_\delta(\widehat{\mathbb{P}}_N)$  defined below, which comprises all finite distributions supported on the finite set  $\widehat{\Xi} := \bigcup_{n \leq N} \widehat{\Xi}_n$ :

$$\begin{aligned} \widehat{\mathbb{B}}_\delta(\widehat{\mathbb{P}}_N) := & \left\{ \{v_\ell\}_{\ell \leq N \cdot 3^{d_\xi}} : \sum_{\ell=1}^{N \cdot 3^{d_\xi}} \sum_{n=1}^N \|\xi_\ell^* - \widehat{\xi}_n\|_1 \pi_{\ell n} \leq \delta, \right. \\ & \sum_{n=1}^N \pi_{\ell n} = v_\ell, \forall \ell \leq N \cdot 3^{d_\xi}, \\ & \sum_{\ell=1}^{N \cdot 3^{d_\xi}} \pi_{\ell n} = 1/N, \forall n \leq N, \\ & \sum_{\ell=1}^{N \cdot 3^{d_\xi}} v_\ell = 1, \\ & v_\ell \geq 0, \forall \ell \leq N \cdot 3^{d_\xi}, \\ & \left. \pi_{\ell n} \geq 0, \forall \ell \leq N \cdot 3^{d_\xi}, \forall n \leq N \right\}. \end{aligned}$$

The distribution separation problem can be formulated as:

$$\max \left\{ \sum_{\ell=1}^{N \cdot 3^{d_\xi}} v_\ell Q(\mathbf{x}, \xi_\ell^*) : \{v_\ell\}_{\ell \leq N \cdot 3^{d_\xi}} \in \widehat{\mathbb{B}}_\delta(\widehat{\mathbb{P}}_N) \right\}. \quad (15)$$

Therefore, we have the following corollary.

**Corollary 3.** *For each  $\mathbf{x} \in X$ , the optimal value of the worst-case expectation problem of the DD-DRO-W problem equals the optimal value of the distribution separation problem which comprises the finite distributions supported on the set  $\widehat{\Xi}$  within Wasserstein ambiguity set.*

*Proof.* We have the following sequence of equivalences:

$$\begin{aligned} & \sup_{\widetilde{P} \in \widehat{\mathbb{B}}_\delta(\widehat{P}_N)} \mathbb{E}^{\widetilde{P}}[Q(\mathbf{x}, \xi)] = \\ & \left\{ \begin{array}{l} \min_{\lambda, s_n} \quad \lambda \delta + \sum_{n=1}^N \frac{1}{N} s_n \\ \text{s.t.} \quad \sup_{\xi \in \widehat{\Xi}} (Q(\mathbf{x}, \xi) - \lambda \|\xi - \widehat{\xi}_n\|_1) \leq s_n, \quad \forall n \leq N, \\ \quad \lambda \geq 0. \end{array} \right. \quad (16) \end{aligned}$$

$$\begin{aligned} & \left\{ \begin{array}{l} \min_{\lambda, s_n} \quad \lambda \delta + \sum_{n=1}^N \frac{1}{N} s_n \\ \text{s.t.} \quad \sup_{\xi \in \widehat{\Xi}_n} (Q(\mathbf{x}, \xi) - \lambda \|\xi - \widehat{\xi}_n\|_1) \leq s_n, \quad \forall n \leq N, \\ \quad \lambda \geq 0. \end{array} \right. \quad (17) \end{aligned}$$

$$= \max \left\{ \sum_{\ell=1}^{N \cdot 3^{d_\xi}} v_\ell Q(\mathbf{x}, \xi_\ell^*) : \{v_\ell\}_{\ell \leq N \cdot 3^{d_\xi}} \in \widehat{\mathbb{B}}_\delta(\widehat{\mathbb{P}}_N) \right\}. \quad (18)$$

Equality (16) is derived by Wiesemann et al. (2014), Zhao and Guan (2018). Equality (17) holds by **Proposition 2**. Finally, equality (18) is obtained by applying again the result from Wiesemann et al. (2014), Zhao and Guan (2018) to the convex reduction of the worst-expectation problem over the Wasserstein ambiguity set for the discrete distributions supported on the set  $\widehat{\Xi}$ .  $\square$

Note that  $|\widehat{\Xi}| = N \cdot 3^{d_\xi}$  by definition of set  $\widehat{\Xi}_n$ , so the decomposition methodology proposed Bansal et al. (2018) may be computationally intractable even for moderately high dimensions. Therefore, for low uncertainty dimensionality  $d_\xi$ , we can address the problem (14) by using the decomposition method presented in Bansal et al. (2018) which uses the distribution separation problem (18). However, as the problem (18) has a large number of variables for high uncertainty dimensionality, it may not be possible to solve it due to memory or time constraints. That precise issue was encountered by Guevara et al. (2020), who circumvented the problem by using machine learning techniques to select the random variables with most impact on the model, and then applying the algorithm of Bansal et al. (2018) only to those variables.

## 4 Decomposition methods

In this section, we present the proposed numerical schemes to solve the tractable reformulation (14) of the two-stage distributionally robust optimization problem with right-hand-side uncertainty under a data-driven Wasserstein based ambiguity set. We show that this problem is suitable to be solved by three exact decomposition methods: the Benders multi-cut and single-cut methods and the column and constraint generation method (C&CG).

In general, a decomposition method can be implemented in a master-oracle scheme; see, e.g., Zeng and Zhao (2013) for an application of that technique to a robust two-stage model. In our context, the master problem is a relaxation of the equivalent linear program (14) of the two-stage DD-DRO-W problem. Given the solution of the master (relaxed) problem, the oracle identifies the worst infeasibility to add the corresponding constraint (or block of constraints and variables) to the master problem. This iterative procedure stops whenever the oracle asserts that the master solution is feasible for the original problem (14).

### 4.1 Column and constraint generation method

We start by proposing a solution methodology to address the problem (14) by using the column and constraint generation method (C&CG). We develop a iterative procedure based on lower and upper bounding approximations of the linear program (14) which converges to its optimal value and optimal solution.

By considering the primal formulation (5) of the recourse function  $Q(\mathbf{x}, \xi)$ , the problem

$$\min_{\mathbf{x}, \mathbf{y}_\ell, \lambda, s_n} \mathbf{c}^\top \mathbf{x} + \lambda \delta + \frac{1}{N} \sum_{n=1}^n s_n \quad (19a)$$

$$\text{s.t.} \quad \mathbf{q}^\top \mathbf{y}_\ell - \lambda \|\xi_\ell^* - \widehat{\xi}_n\|_1 \leq s_n, \quad \forall \ell \in \mathcal{L}_n^K, \forall n \leq N, \quad (19b)$$

$$\mathbf{W} \mathbf{y}_\ell = \mathbf{H}(\mathbf{x}) \xi_\ell^* + \mathbf{r}(\mathbf{x}), \quad \forall \ell \in \mathcal{L}_n^K, \forall n \leq N, \quad (19c)$$

$$\mathbf{y}_\ell \geq 0, \quad \forall \ell \in \mathcal{L}_n^K, \forall n \leq N, \quad (19d)$$

$$\lambda \geq 0, \quad (19e)$$

$$\mathbf{x} \in X, \quad (19f)$$

is equivalent to (14) if the set  $\mathcal{L}_n^K$  equals  $\mathcal{L}_n$ . Instead, if  $\mathcal{L}_n^K \subset \mathcal{L}_n$  is a subset defined at iteration  $K$  of the iterative procedure, problem (19) represents a relaxation of (14). Therefore, its optimal value is a valid lower bound,  $LB$ , for the optimal value of problem (14). Henceforth, we call the relaxed problem (19) as the master problem.

For the current optimal solution  $(\mathbf{x}^K, \lambda^K, \mathbf{s}^K)$  of the master problem (19), where  $\mathbf{s}^K = (s_n^K)_{n \leq N}$ , we need an oracle problem to find the worst-case uncertainty realization  $\xi_\ell^* \in \widehat{\Xi}_n$  maximizing the infeasibility of constraint (19b). Based on this result, we update the subset  $\mathcal{L}_n^K$ , for all  $n = 1, \dots, N$ , and add the block of linear constraints (19b)-(19c) and variables (19d) - new columns - to the master problem. With this in



mind, let us consider the left-hand-side of the inequality (7). We have the following equivalences:

$$\sup_{\xi \in \Xi} (Q(\mathbf{x}^K, \xi) - \lambda^K \|\xi - \widehat{\xi}_n\|_1) \quad (20a)$$

$$= \left\{ \begin{array}{l} \sup_{\xi \in \Xi} \min_{\mathbf{y} \geq 0} \mathbf{q}^\top \mathbf{y} - \lambda^K \|\xi - \widehat{\xi}_n\|_1 \\ \text{s.t. } \mathbf{W}\mathbf{y} = \mathbf{H}(\mathbf{x}^K)\xi + \mathbf{r}(\mathbf{x}^K) \end{array} \right\} \quad (20b)$$

$$= \left\{ \begin{array}{l} \max_{\theta, \xi} \theta^\top \mathbf{r}(\mathbf{x}^K) + [\mathbf{H}^\top(\mathbf{x}^K)\theta]^\top \xi - \lambda^K \|\xi - \widehat{\xi}_n\|_1 \\ \text{s.t. } \mathbf{W}^\top \theta \leq \mathbf{q} \\ \xi \in \Xi. \end{array} \right\} \quad (20c)$$

We consider the bilinear problem (20c) as the oracle problem which can be reduced to a MILP problem by linearizing the products of binary and continuous variables. To that end, let us consider the inner problem in the left-hand side of inequality (7) and introduce the notation  $\Delta^+ = \mathbf{b} - \widehat{\xi}_n$ ,  $\Delta^- = \widehat{\xi}_n - \mathbf{a}^3$ . With this notation at hand, according to **Proposition 2**, the optimal solution  $\xi^*$  of problem (8) - which is equivalent to problem (20c) - can be expressed by:

$$\xi^* = \widehat{\xi}_n + \text{diag}(\Delta^+) \mathbf{z}^+ - \text{diag}(\Delta^-) \mathbf{z}^-,$$

where  $\mathbf{z}^+, \mathbf{z}^- \in \{0, 1\}^{d_\xi}$  are binary vector variables and  $\text{diag}(\Delta^i) = \text{diag}(\Delta_1^i, \dots, \Delta_{d_\xi}^i)$ , denotes the diagonal matrix of the vector  $\Delta^i$ , for  $i = +, -$ , i.e.,

$$\begin{aligned} \xi^* &= \begin{bmatrix} \widehat{\xi}_1 \\ \vdots \\ \widehat{\xi}_{d_\xi} \end{bmatrix} + \begin{bmatrix} \Delta_1^+ & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \Delta_{d_\xi}^+ \end{bmatrix} \begin{bmatrix} z_1^+ \\ \vdots \\ z_{d_\xi}^+ \end{bmatrix} - \begin{bmatrix} \Delta_1^- & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \Delta_{d_\xi}^- \end{bmatrix} \begin{bmatrix} z_1^- \\ \vdots \\ z_{d_\xi}^- \end{bmatrix} \\ &= \begin{bmatrix} \widehat{\xi}_1 + \Delta_1^+ z_1^+ - \Delta_1^- z_1^- \\ \vdots \\ \widehat{\xi}_{d_\xi} + \Delta_{d_\xi}^+ z_{d_\xi}^+ - \Delta_{d_\xi}^- z_{d_\xi}^- \end{bmatrix}. \end{aligned}$$

Thus, the decision variable  $\xi$  of the optimization problem (8) can be replaced by binary decision variables  $\mathbf{z}^i$ ,  $i = +, -$ , and the optimal value of problem (8) equals:

$$\begin{aligned} \max_{\theta, \mathbf{z}^+, \mathbf{z}^-} \quad & \theta^\top \mathbf{r}(\mathbf{x}^K) + [\mathbf{H}^\top(\mathbf{x}^K)\theta]^\top [\widehat{\xi}_n + \text{diag}(\Delta^+) \mathbf{z}^+ - \text{diag}(\Delta^-) \mathbf{z}^-] - \lambda^K \mathbf{1}^\top \boldsymbol{\alpha} \\ \text{s.t.} \quad & \mathbf{W}^\top \theta \leq \mathbf{q} \\ & \widehat{\xi}_n + \text{diag}(\Delta^+) \mathbf{z}^+ - \text{diag}(\Delta^-) \mathbf{z}^- \leq \boldsymbol{\alpha} \\ & \widehat{\xi}_n + \text{diag}(\Delta^-) \mathbf{z}^- - \text{diag}(\Delta^+) \mathbf{z}^+ \leq \boldsymbol{\alpha}. \end{aligned} \quad (21)$$

However, problem (21) has bilinear terms of products of binary and continuous variables which can be linearized by disjunctive constraint following Fortuny-Amat and McCarl (1981). For the detailed exact linearized mixed integer linear programming (MILP) formulation, see Appendix A.

Let us denote by  $\rho_n(\mathbf{x}^K, \lambda^K)$  the optimal value of the oracle problem (21) (or equivalently problem (20c)). For a fixed  $(\mathbf{x}^K, \lambda^K)$  we claim that

$$s_n^K \leq \rho_n(\mathbf{x}^K, \lambda^K), \quad \forall n \leq N. \quad (22)$$

Indeed, we have that constraint (19b) is active for the  $\ell \in \mathcal{L}_n^K$  that maximizes the left-hand-side of constraint (19b) over the set  $\mathcal{L}_n^K$ , for all  $n = 1, \dots, N$ , i.e.,

$$\max_{\ell \in \mathcal{L}_n^K} (\mathbf{q}^\top \mathbf{y}_\ell^* - \lambda^K \|\xi_\ell^* - \widehat{\xi}_n\|_1) = s_n^K, \quad \forall n \leq N, \quad (23)$$

where  $\mathbf{y}_\ell^*$  denotes the optimal solution of the primal second-stage variable  $\mathbf{y}_\ell$ , for all  $\ell \in \mathcal{L}_n^K$ ,  $n = 1, \dots, N$ , within problem (19). Therefore, we have the following valid inequality:

$$\rho_n(\mathbf{x}^K, \lambda^K) = \sup_{\xi \in \Xi} \left( Q(\mathbf{x}^K, \xi) - \lambda^K \|\xi - \widehat{\xi}_n\|_1 \right) \quad (24a)$$

$$\geq \max_{\ell \in \mathcal{L}_n^K} \left( \mathbf{q}^\top \mathbf{y}_\ell^* - \lambda^K \|\xi_\ell^* - \widehat{\xi}_n\|_1 \right) = s_n^K, \quad (24b)$$

for all  $n = 1, \dots, N$ . Equality (24a) holds because the equivalence between problems (20c) and (20a), whereas (24b) follows from the fact that the optimization problem on the right-hand side of (24a) is less constrained than that of (24b). Hence, a valid upper bound  $UB$  for the problem (14) can be obtained as

$$UB = \mathbf{c}^\top \mathbf{x}^K + \lambda^K \delta + \frac{1}{N} \sum_{n=1}^N \rho_n(\mathbf{x}^K, \lambda^K). \quad (25)$$

The algorithm converges whenever the  $UB - LB \leq \varepsilon$ , i.e, the current solution  $(\mathbf{x}^K, \lambda^K, \mathbf{s}_n^K)$  lies within a user-defined tolerance level  $\varepsilon$ .

We summarize the column and constraint generation algorithm in the following pseudo-code:

---

**Algorithm 1** Column and constraint generation method

---

**Initialization:** Set  $K = 0$ ,  $UB \leftarrow +\infty$  and  $LB \leftarrow -\infty$ , and  $\mathcal{L}_n^K \leftarrow \{n\}$  for  $n = 1, \dots, N$   
**while**  $UB - LB > \varepsilon$  **do**  
    Solve the Master problem (19);  
    Store the Master solution:  $(LB, \mathbf{x}^K, \lambda^K, \mathbf{s}^K)$ ;  
    **for**  $n=1$  **to**  $N$  **do**  
        Solve the MILP version of the Oracle problem (21) for  $(\mathbf{x}^K, \lambda^K)$ ;  
        Store the Oracle solution:  $\rho_n(\mathbf{x}^K, \lambda^K), (\theta_n^*, \xi_n^*)$ ;  
        **if**  $\rho_n(\mathbf{x}^K, \lambda^K) > s_n^K$  **then**  
             $\ell \leftarrow N(K+1) + n$ ;  
            Update  $\mathcal{L}_n^{K+1} \leftarrow \mathcal{L}_n^K \cup \{\ell\}$  and make  $\xi_\ell^* \leftarrow \xi_n^*$   
        **end if**  
    **end for**  
    Set  $UB \leftarrow \min\{UB, \mathbf{c}^\top \mathbf{x}^K + \lambda^K \delta + \frac{1}{N} \sum_{n=1}^N \rho_n(\mathbf{x}^K, \lambda^K)\}$   
    Update  $K \leftarrow K + 1$ , add the block of linear constraints (19b)–(19c) and variables (19d)  
**end while**  
**return**  $\mathbf{x}^K, UB, LB$

---

For initialization purpose, we can solve the second-stage problem

$$\min_{\mathbf{y} \geq 0} \mathbf{q}^\top \mathbf{y} \quad (26a)$$

$$\text{s.t. } \mathbf{W}\mathbf{y} = \mathbf{H}(\bar{\mathbf{x}})\widehat{\xi}_n + \mathbf{r}(\bar{\mathbf{x}}) \quad : \theta_n \quad (26b)$$

for the optimal solution  $\bar{\mathbf{x}}$  of the deterministic equivalent problem for the average scenario,  $\bar{\xi} = \frac{1}{N} \sum_{n=1}^N \widehat{\xi}_n$ , and cast the dual variable  $\theta_n$  of the constraint (26b)—which is a dual vertex—for all  $n = 1, \dots, N$ . We then initialize the algorithm (1) with  $\xi_n^* \leftarrow \widehat{\xi}_n$  for all  $n = 1, \dots, N$ .

## 4.2 Multi-cut Benders

We can also address the problem (14) by using a multi-cut Benders algorithm. By strong duality, we can assess  $Q(\mathbf{x}, \xi_\ell^*)$  by:

$$Q(\mathbf{x}, \xi_\ell^*) = \max_{d \in \mathcal{D}} \left\{ \theta_d^\top \mathbf{r}(\mathbf{x}) + [\mathbf{H}^\top(\mathbf{x})\theta_d]^\top \xi_\ell^* \right\}, \quad (27)$$

for all  $\ell \in \mathcal{L}_n$  and  $\mathbf{x} \in X$ , where  $\{\theta_d\}_{d \in \mathcal{D}}$  is the set of vertices of the dual polyhedron  $\{\theta : \mathbf{W}^\top \theta \leq \mathbf{q}\}$ , hereafter referred to as dual vertices. Thus, by replacing  $Q(\mathbf{x}, \xi_\ell^*)$  with the enumeration of the affine functions  $\{\theta_d^\top \mathbf{r}(\mathbf{x}) + [\mathbf{H}^\top(\mathbf{x})\theta_d]^\top \xi_\ell^*\}_{d \in \mathcal{D}}$  in (14b), the problem

$$\min_{\mathbf{x}, \lambda, s_n} \mathbf{c}^\top \mathbf{x} + \lambda \delta + \frac{1}{N} \sum_{n=1}^N s_n \quad (28a)$$

$$\text{s.t.} \quad \theta_d^\top \mathbf{r}(\mathbf{x}) + [\mathbf{H}^\top(\mathbf{x})\theta_d]^\top \xi_\ell^* - \lambda \|\xi_\ell^* - \widehat{\xi}_n\|_1 \leq s_n, \quad (28b)$$

$$\forall (d, \ell) \in \mathcal{D} \times \mathcal{L}_n, \forall n \leq N,$$

$$\lambda \geq 0, \quad (28c)$$

$$\mathbf{x} \in X, \quad (28d)$$

meets problem (14).

We can derive an alternative master problem from the linear programming relaxation of the equivalent problem (28). Observe that by solving the oracle problem (20c), we obtain an uncertainty realization and a dual vertex that maximizes the infeasibility of constraint (28b). Let

$$(\theta_n^k, \xi_n^k) \in \arg \max_{\theta, \xi} \left\{ \theta^\top \mathbf{r}(\mathbf{x}^k) + [\mathbf{H}^\top(\mathbf{x}^k)\theta]^\top \xi - \lambda^k \|\xi - \widehat{\xi}_n\|_1 \mid \begin{array}{l} \mathbf{W}^\top \theta \leq \mathbf{q}, \\ \xi \in \Xi \end{array} \right\} \quad (29)$$

denote the optimal solution of the oracle problem, where  $(\mathbf{x}^k, \lambda^k)$  is the optimal solution, at iteration  $k \leq K$ , of the master problem for the finite extensive equivalent form (28). Therefore, a linear programming relaxation of (28) can be derived by substituting constraint (28b) with optimality cuts:

$$[\theta_n^k]^\top \mathbf{r}(\mathbf{x}) + [\mathbf{H}^\top(\mathbf{x})\theta_n^k]^\top \xi_n^k - \lambda \|\xi_n^k - \widehat{\xi}_n\|_1 \leq s_n, \forall k \leq K, n \leq N, \quad (30)$$

Note that for  $K$  sufficiently large, the relaxed problem equals (28) if

$$\{(\theta_n^k, \xi_n^k) \mid k \leq K\} = \{(\theta_d, \xi_\ell^*) \mid d \in \mathcal{D}, \ell \in \mathcal{L}_n\},$$

for all  $n = 1, \dots, N$ .

### 4.3 Single-cut Benders

For the purpose of constructing the single-cut Benders algorithm, we develop an equivalent formulation that incorporate additional valid constraints to construct the average cut. The equivalent formulation for (28) is obtained by replacing (28b) with

$$\theta_d^\top \mathbf{r}(\mathbf{x}) + [\mathbf{H}^\top(\mathbf{x})\theta_d]^\top \xi_\ell^* - \lambda \|\xi_\ell^* - \widehat{\xi}_n\|_1 \leq s_n, \forall (d, \ell) \in \mathcal{D} \times \mathcal{L}, \forall n \leq N, \quad (31)$$

where  $\mathcal{L} = \cup_{n \leq N} \mathcal{L}_n$ . To see that such equivalence holds, notice that

$$\max_{\ell \in \mathcal{L}} Q(\mathbf{x}, \xi_\ell^*) - \lambda \|\xi_\ell^* - \widehat{\xi}_n\|_1 \leq \sup_{\xi \in \Xi} Q(\mathbf{x}, \xi) - \lambda \|\xi - \widehat{\xi}_n\|_1 \quad (32a)$$

$$= \max_{\ell \in \mathcal{L}_n} Q(\mathbf{x}, \xi_\ell^*) - \lambda \|\xi_\ell^* - \widehat{\xi}_n\|_1 \quad (32b)$$

$$\leq \max_{\ell \in \mathcal{L}} Q(\mathbf{x}, \xi_\ell^*) - \lambda \|\xi_\ell^* - \widehat{\xi}_n\|_1. \quad (32c)$$

The first inequality (32a) is valid since  $\{\xi_\ell\}_{\ell \in \mathcal{L}} \subseteq \Xi$ , while the second equality (32b) is guaranteed by Proposition 2. Finally, the last inequality (32c) holds since  $\{\xi_\ell\}_{\ell \in \mathcal{L}_n} \subseteq \{\xi_\ell\}_{\ell \in \mathcal{L}}$ . It follows that

$$\max_{\ell \in \mathcal{L}} Q(\mathbf{x}, \xi_\ell^*) - \lambda \|\xi_\ell^* - \widehat{\xi}_n\|_1 = \max_{\ell \in \mathcal{L}_n} Q(\mathbf{x}, \xi_\ell^*) - \lambda \|\xi_\ell^* - \widehat{\xi}_n\|_1$$

and consequently

$$\begin{aligned} & \max_{(d,\ell) \in \mathcal{D} \times \mathcal{L}} \boldsymbol{\theta}_d^\top \mathbf{r}(\mathbf{x}) + [\mathbf{H}^\top(\mathbf{x})\boldsymbol{\theta}_d]^\top \boldsymbol{\xi}_\ell^* - \lambda \|\boldsymbol{\xi}_\ell^* - \widehat{\boldsymbol{\xi}}_n\|_1 \\ & = \max_{(d,\ell) \in \mathcal{D} \times \mathcal{L}_n} \boldsymbol{\theta}_d^\top \mathbf{r}(\mathbf{x}) + [\mathbf{H}^\top(\mathbf{x})\boldsymbol{\theta}_d]^\top \boldsymbol{\xi}_\ell^* - \lambda \|\boldsymbol{\xi}_\ell^* - \widehat{\boldsymbol{\xi}}_n\|_1. \end{aligned}$$

Now, we can rewrite the problem (28) as a finite (extensive) linear program with average cuts by replacing (28b) with (31) and taking the average of the latter over  $n$ :

$$\min_{\mathbf{x}, \lambda, \beta} \mathbf{c}^\top \mathbf{x} + \lambda \delta + \beta \tag{33a}$$

$$\text{s.t. } \frac{1}{N} \sum_{n=1}^N \left( [\boldsymbol{\theta}_d]^\top \mathbf{r}(\mathbf{x}) + [\mathbf{H}^\top(\mathbf{x})\boldsymbol{\theta}_d]^\top \boldsymbol{\xi}_\ell - \lambda \|\boldsymbol{\xi}_\ell - \widehat{\boldsymbol{\xi}}_n\|_1 \right) \leq \beta, \forall (d, \ell) \in \mathcal{D} \times \mathcal{L}, \tag{33b}$$

$$\lambda \geq 0, \tag{33c}$$

$$\mathbf{x} \in X. \tag{33d}$$

Accordingly, we derive an alternative master-oracle scheme that uses the same oracle (29), but with the modified master

$$\min_{\mathbf{x}, \lambda, \beta} \mathbf{c}^\top \mathbf{x} + \lambda \delta + \beta \tag{34a}$$

$$\text{s.t. } \frac{1}{N} \sum_{n=1}^N \left( [\boldsymbol{\theta}_n^k]^\top \mathbf{r}(\mathbf{x}) + [\mathbf{H}^\top(\mathbf{x})\boldsymbol{\theta}_n^k]^\top \boldsymbol{\xi}_n^k - \lambda \|\boldsymbol{\xi}_n^k - \widehat{\boldsymbol{\xi}}_n\|_1 \right) \leq \beta, \forall k \leq K, \tag{34b}$$

$$\lambda \geq 0, \tag{34c}$$

$$\mathbf{x} \in X, \tag{34d}$$

where  $(\boldsymbol{\theta}_n^k, \boldsymbol{\xi}_n^k)$  in this case denote the optimal solution of the oracle (29) for the optimal solution  $(\mathbf{x}^k, \lambda^k)$  of the problem (34) at iteration  $k - 1$ . Note that the resulting master problem (34) is less constrained than the linear programming relaxation (28). Therefore, the multi-cut algorithm provides a tighter lower bound than the single-cut master problem (34). In turn, the single-cut version of the algorithm relies on a reduced number of constraints (cuts). While this implies a lower computational effort to the master problem, it is very likely that the multi-cut will converge in fewer iterations. Thus, there is a tradeoff between these two versions of the algorithm that should be empirically studied, as we do in the next section.

## 5 Numerical experiments

In this section, we test the proposed decomposition methods based on a well-known energy application, the unit commitment problem with net load uncertainty. First, we present the problem and then we analyze the computational efficiency of each proposed decomposition method. We benchmark results with existing algorithms.

### 5.1 Data-driven unit commitment formulation

As discussed in Zhao and Guan (2018), when the probability distribution of the electricity load can not be accurately estimated, the obtained unit commitment decision can be biased. In the literature, DRO models have been already been applied to the unit commitment problem (see Zhu et al. (2019) and references therein). However, to the best of our knowledge, existing works rely mainly on approximations, such as affine policies, for the second stage variables. Our paper aims to address precisely this issue.

For illustrative purposes, we use a classic and simple formulation of a data-driven two-stage stochastic single-bus unit commitment problem (Papavasiliou and Oren 2013). The first stage comprises the

commitment decisions while the second stage accounts for the dispatch decisions. We develop the problem by minimizing the expected total generation cost, and the uncertainty on the right-hand-side of the problem in the net electricity load parameter (load subtracted by uncertain renewable injections).

We summarize the notations by sets, parameters, first-stage variables and second-stage variables listed as follows:

Set	Description
$\mathcal{I}$	Set of electricity generators
$\mathcal{T}$	Set of time of periods
$\mathcal{S}$	Set of electricity load scenarios

Table 1: Description of the sets.

Parameter	Description
$C_i^u$	Fixed cost of unit $i \in \mathcal{I}$
$C_i^{SU}$	Start-up cost of unit $i \in \mathcal{I}$
$C_i^{SD}$	Shut-down cost of unit $i \in \mathcal{I}$
$R_i^U$	Ramp-up limit of unit $i \in \mathcal{I}$
$R_i^D$	Ramp-down limit of unit $i \in \mathcal{I}$
$\overline{P}_i$	Maximum power generation of unit $i \in \mathcal{I}$
$\underline{P}_i$	Minimum power generation of unit $i \in \mathcal{I}$
$\widehat{\xi}_n(t)$	The electricity load in time $t \in \mathcal{T}$ corresponding to scenario $n \leq N$

Table 2: Description of the parameters.

Variable	Description
$u_{i,t}$	Binary commit variable: 1 if the thermal generator $i$ is on in time $t$ ; 0 otherwise
$v_{i,t}$	Start-up variable for unit $i \in \mathcal{I}$ in time $t \in \mathcal{T}$
$w_{i,t}$	Shut-down variable for unit $i \in \mathcal{I}$ in time $t \in \mathcal{T}$

Table 3: Description of the first-stage variables.

Variable	Description
$p_{i,t}$	Power generation of unit $i \in \mathcal{I}$ in time $t \in \mathcal{T}$

Table 4: Description of the second-stage variables.

Based on this notation, the mathematical formulation of the data-driven two-stage stochastic unit commitment problem is as follows:

$$\min_{u_{i,t}, v_{i,t}, w_{i,t}, p_{i,t}} \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \left[ C_i^u u_{i,t} + C_i^{SU} v_{i,t} + C_i^{SD} w_{i,t} + \frac{1}{N} \sum_{n=1}^N [C_i^P p_{i,t}^n] \right] \quad (35a)$$

$$s.t. \quad \sum_{i \in \mathcal{I}} p_{i,t}^n = \widehat{\xi}_n(t), \quad \forall t \in \mathcal{T}, \forall n \leq N, \quad (35b)$$

$$v_{i,t} - w_{i,t} = u_{i,t} - u_{i,t-1}, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \quad (35c)$$

$$v_{i,t} \leq u_{i,t}, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \quad (35d)$$

$$w_{i,t} \leq 1 - u_{i,t}, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \quad (35e)$$

$$p_{i,t} - p_{i,t-1} \leq R_i^U u_{i,t-1} + \bar{P}_i v_{i,t}, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \quad (35f)$$

$$p_{i,t-1} - p_{i,t} \leq R_i^D u_{i,t} + \underline{P}_i w_{i,t}, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \quad (35g)$$

$$\underline{P}_i u_{i,t} \leq p_{i,t} \leq \bar{P}_i u_{i,t}, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \quad (35h)$$

$$0 \leq v_{i,t} \leq 1, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \quad (35i)$$

$$0 \leq w_{i,t} \leq 1, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \quad (35j)$$

$$u_{i,t} \in \{0, 1\}, \quad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}. \quad (35k)$$

In the above formulation, constraint (35b) ensures load balance. Constraints (35c)-(35d) are the start-up and shut-down operational constraints for each thermal unit. Constraints (35e) and (35f) are the ramping up and ramping down constraints, respectively. Finally, constraint (35g) is the minimum and maximum power generation of the unit  $i \in \mathcal{I}$  in time  $t \in \mathcal{T}$ .

We consider the distributionally robust optimization version of the problem (35). Regarding to the empirical distribution of historical data, we construct the Wasserstein-based ambiguity set for a given confidence level  $\delta > 0$ . We then consider the formulation (4) of the problem (35), where the first-stage variable  $\mathbf{x}$  is  $[u_{i,t}, v_{i,t}, w_{i,t}]_{i \in \mathcal{I}, t \in \mathcal{T}}$  and  $Q(\mathbf{x}, \xi)$  represents the economic dispatch problem.

## 5.2 Results

For performance comparison purposes, we solve a single-bus system with  $I = 5, 14, 54$  thermal generators over a 24-hour operational time. We develop scenarios for the electricity load over 24-hour span time by setting a deterministic profile distribution of the load and discounting wind power generation. The source of data for the wind power generation is the Global Energy Competition (GEFCom) (Hong et al. 2014, 2016). For reproducibility purposes, the system data is presented in Appendix B.

With the scenarios at hand, we construct the Wasserstein ball around the empirical distribution. Although the number of training samples  $N$  can be estimated for a given confidence level  $\delta > 0$ , for illustration purposes in our computational experiments we consider a fixed number of training samples equal to 100 and a fixed parameter  $\delta$  equal to 3.

The computational experiments were implemented using JuMP (Dunning et al. 2017), a modeling language for mathematical optimization embedded in the Julia programming language. The solver Gurobi 7.5.2 was used as the MIP solver to run the computational experiments on an Intel Core i7, a 4.0-GHz processor with 32 GB of RAM.

For illustrative purposes, we report in Table 5 the gap—the difference between the upper and lower bounding approximation—at iteration  $K$  and the computing time of each algorithm, for the system with 5 thermal generators. We used  $\varepsilon = 0.0$ . The C&CG method is clearly superior. The results show that the oracle sub problem asserts feasibility faster for the current solution of the master problem  $(\mathbf{x}^K, \lambda^K)$  of the C&CG method. This is the main advantage of this algorithm, whenever the computational burden of the master problem can be dealt with.

<b>Iteration</b>	<b>C&amp;CG</b>	<b>Benders multi-cut</b>	<b>Benders single-cut</b>
<i>K</i>	<i>UB - LB</i>	<i>UB - LB</i>	<i>UB - LB</i>
1	1273245	1404389	1404388
2	0	147243	282048
3		135559	135559
4		8471	48166
5		8064	19806
6		10656	19806
7		1670	19490
8		1274	16295
9		5548	16295
10		756	16295
11		1278	16295
12		2946	16295
13		760	15564
14		799	15564
15		863	15041
16		799	14310
17		0	13777
40			9142
80			3857
160			2218
165			0
<b>Time (CPUs)</b>	3394	26622	148127

Table 5: Data-driven Unit Commitment problem with 5 generators.

Table 6 reports the computing time of convergence of each algorithm for the systems with 14 and 54 thermal generators, respectively. Symbol (-) means that there is no computing time to report. For the system with 14 thermal generators, all algorithms converged in moderate computing time, whereas for the system with 54 thermal generators, only the C&CG algorithm converged, which confirm the superiority of this last method even for these large systems.

Method	Syst-14	Syst-54
	Time (CPUs)	Time (CPUs)
C&CG	15365	24225
Bender's multi-cut	27833	-
Bender's single-cut	38379	-

Table 6: Computing time.

We also benchmark the proposed solution methodology with the existing solution approaches applied to the class of parametric linear programs with right-hand-side uncertainty. First, we consider the tractable reformulation of the TS-DRO-W problem derived from the convex reduction develop in Mohajerin Esfahani and Kuhn (2018) which scales with the number of dual vertices. We use the algorithm for enumerating vertices embedding in the computational tool Polyhedral.jl of the programming language Julia. However, the enumerating algorithm does not converge in reasonable time for the considered instances.

Given the equivalence of the TS-DRO-W problem with the TS-DRO problem under the Wasserstein ambiguity set which comprises the finite distributions supported on the set  $\widehat{\Xi}$ , we utilize the algorithmic decomposition method presented in Bansal et al. (2018) for two-stage distributionally robust mixed binary programs to the address unit commitment instances. Nevertheless, the resulting distribution separation problem could not be solved, because the enumeration of the set  $\widehat{\Xi}$  has  $N \cdot 3^{d_\xi}$  complexity which is out-of-memory for the considered uncertainty dimensionality.

## 6 Conclusion

In this paper we have presented a new algorithmic approach to solve distributionally robust optimization problems with right-hand-side uncertainty over Wasserstein balls. Our model assumes distributions supported on an infinite compact set within a Wasserstein ball, which is built around the empirical distribution of observed data of the uncertainty. The proposed approach and formulation allow us solving the extensive form of a convex reformulation of the problem very efficiently.

This issue is paramount of importance because of the existing algorithmic schemes are computationally intractable for instances with high uncertain dimensionality or an exponential number of dual vertices. Instead, we have proposed a finite extensive equivalent form of the problem which is solved by using an exact decomposition algorithm that converges in a finite number of iterations.

We tested the proposed algorithms with a two-stage unit commitment problem for the day-ahead scheduling of power generation by assuming an uncertain energy load. We analyzed the computational performance of each algorithm by varying the size of the considered systems. In particular, there are no tractable formulations in the literature for these instances, which demonstrates that our proposed approach is a substantial advance to address the class of DRO problems with right-hand side uncertainty.

The results show that when the second-stage problem has a complex polyhedral structure, the C&CG has the best computational performance among the three methods that were tested. The superiority of C&CG stems from the fact that it optimizes the actual value of the second-stage function for each



uncertainty realization, unlike the approaches based on Benders methods that optimize only a piecewise linear approximation of this function.

## Appendix A: Exact separation approach for the oracle

For a given solution  $(\mathbf{x}^K, \lambda^K)$ , by introducing auxiliary variables  $\bar{\mathbf{w}}^j = z_j^+ \boldsymbol{\theta}$  and  $\underline{\mathbf{w}}^j = z_j^- \boldsymbol{\theta}$ , the objective function of problem (8) is equivalent to

$$\begin{aligned}
& \boldsymbol{\theta}^\top \mathbf{r}(\mathbf{x}^K) + [\mathbf{H}^\top(\mathbf{x}^K) \boldsymbol{\theta}]^\top [\widehat{\boldsymbol{\xi}}_n + \text{diag}(\Delta^+) \mathbf{z}^+ - \text{diag}(\Delta^-) \mathbf{z}^-] \\
& - \lambda^K \mathbf{1}^\top \boldsymbol{\alpha} \\
= & \sum_{i=1}^{m_y} \theta_i r_i(\mathbf{x}^K) + \sum_{j=1}^{d_\xi} \sum_{i=1}^{m_y} H_{j,i}(\mathbf{x}^K) \theta_i (\widehat{\boldsymbol{\xi}}_n)_j + \sum_{j=1}^{d_\xi} \sum_{i=1}^{m_y} H_{j,i}(\mathbf{x}^K) \underbrace{\theta_i z_j^+}_{\bar{w}_i^j} \Delta_j^+ \\
& - \sum_{j=1}^{d_\xi} \sum_{i=1}^{m_y} H_{j,i}(\mathbf{x}^K) \underbrace{\theta_i z_j^-}_{\underline{w}_i^j} \Delta_j^- - \lambda^K \sum_{j=1}^{d_\xi} \alpha_j \\
= & \sum_{i=1}^{m_y} \theta_i r_i(\mathbf{x}^K) + \sum_{j=1}^{d_\xi} \sum_{i=1}^{m_y} H_{j,i}(\mathbf{x}^K) \theta_i (\widehat{\boldsymbol{\xi}}_n)_j + \sum_{j=1}^{d_\xi} \sum_{i=1}^{m_y} H_{j,i}(\mathbf{x}^K) \bar{w}_i^j \Delta_j^+ \\
& - \sum_{j=1}^{d_\xi} \sum_{i=1}^{m_y} H_{j,i}(\mathbf{x}^K) \underline{w}_i^j \Delta_j^- - \lambda^K \sum_{j=1}^{d_\xi} \alpha_j.
\end{aligned}$$

Then, the oracle (8) has a MILP equivalent given by:

$$\begin{aligned}
& \max_{\boldsymbol{\theta}, \mathbf{z}^+, \bar{\mathbf{w}}^j, \mathbf{z}^-, \underline{\mathbf{w}}^j, \boldsymbol{\alpha}} \left\{ \begin{aligned} & \sum_{i=1}^{m_y} \theta_i r_i(\mathbf{x}^K) + \sum_{j=1}^{d_\xi} \sum_{i=1}^{m_y} H_{j,i}(\mathbf{x}^K) \theta_i (\widehat{\boldsymbol{\xi}}_n)_j + \sum_{j=1}^{d_\xi} \sum_{i=1}^{m_y} H_{j,i}(\mathbf{x}^K) \bar{w}_i^j \Delta_j^+ \\ & - \sum_{j=1}^{d_\xi} \sum_{i=1}^{m_y} H_{j,i}(\mathbf{x}^K) \underline{w}_i^j \Delta_j^- - \lambda^K \sum_{j=1}^{d_\xi} \alpha_j \end{aligned} \right\} \\
\text{s.t.} \quad & \sum_{i=1}^{m_y} W_{ji} \theta_i \leq q_j, \quad j = 1, \dots, d_y, \\
& (\widehat{\boldsymbol{\xi}}_n)_j + \Delta_j^+ z_j^+ - \Delta_j^- z_j^- \leq \alpha_j, \quad j = 1, \dots, d_\xi, \\
& (\widehat{\boldsymbol{\xi}}_n)_j + \Delta_j^- z_j^- - \Delta_j^+ z_j^+ \leq \alpha_j, \quad j = 1, \dots, d_\xi, \\
& |\bar{w}_i^j - \theta_i| \leq (1 - z_j^+) M \quad j = 1, \dots, d_\xi, \quad i = 1, \dots, m_y, \\
& |\underline{w}_i^j - \theta_i| \leq (1 - z_j^-) M \quad j = 1, \dots, d_\xi, \quad i = 1, \dots, m_y, \\
& |\bar{w}_i^j| \leq M z_j^+, \quad j = 1, \dots, d_\xi, \quad i = 1, \dots, m_y, \\
& |\underline{w}_i^j| \leq M z_j^-, \quad j = 1, \dots, d_\xi, \quad i = 1, \dots, m_y, \\
& z_j^+ + z_j^- \leq 1 \quad j = 1, \dots, d_\xi, \\
& z_j^+, z_j^- \in \{0, 1\}, \quad j = 1, \dots, d_\xi,
\end{aligned} \tag{36}$$

where  $M \in \mathbb{R}$  is sufficiently large and the products  $\theta_i z_j^+$  and  $\theta_i z_j^-$ , for  $j = 1, \dots, d_\xi$ ,  $i = 1, \dots, m_y$ , are linearized.

## Appendix B: Data

Unit	$C^u$ (\$)	$C^P$ (\$/MW)	$C^{SU}$ (\$)	$C^{SD}$ (\$)	$R^U$ (MW)	$R^D$ (MW)	$\underline{P}$ (MW)	$\bar{P}$ (MW)
1	50	10	100	100	5	5	50	80
2	50	20	100	100	15	15	40	70
3	50	50	50	50	25	50	10	60
4	50	200	10	10	50	50	5	60
5	50	300	10	10	60	60	1	60

Table 7: Parameters of the single-bus system with 5 thermal generators.

## References

- Manish Bansal, Kuo-Ling Huang, and Sanjay Mehrotra. Decomposition algorithms for two-stage distributionally robust mixed binary programs. *SIAM Journal on Optimization*, 28(3):2360–2383, 2018. doi: 10.1137/17M1115046. URL <https://doi.org/10.1137/17M1115046>.
- Rüdiger Barth, Heike Brand, Peter Meibom, and Christoph Weber. A stochastic unit-commitment model for the evaluation of the impacts of integration of large amounts of intermittent wind power. In *Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on*, pages 1–8. IEEE, 2006.
- I. Dunning, J. Huchette, and M. Lubin. Jump: A modeling language for mathematical optimization. *SIAM Review*, 59(2):295–320, 2017. doi: 10.1137/15M1020575. URL <https://doi.org/10.1137/15M1020575>.
- Daniel Duque, Sanjay Mehrotra, and David P. Morton. Distributionally robust two-stage stochastic programming. *Optimization online*, 2020.
- José Fortuny-Amat and Bruce McCarl. A representation and economic interpretation of a two-level programming problem. *Journal of the operational Research Society*, 32(9):783–792, 1981.
- Komei Fukuda and Alain Prodon. Double description method revisited. In *Franco-Japanese and Franco-Chinese Conference on Combinatorics and Computer Science*, pages 91–111. Springer, 1995.
- E. Guevara, F. Babonneau, T. Homem-de-Mello, and S. Moret. A machine learning and distributionally robust optimization framework for strategic energy planning under uncertainty. *Applied Energy*, 271:115005, 2020.
- Grani A. Hanasusanto and Daniel Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Operations Research*, 66(3):849–869, 2018. doi: 10.1287/opre.2017.1698. URL <https://doi.org/10.1287/opre.2017.1698>.
- Tao Hong, Pierre Pinson, and Shu Fan. Global energy forecasting competition 2012. *International Journal of Forecasting*, 30(2):357 – 363, 2014. ISSN 0169-2070. doi: <https://doi.org/10.1016/j.ijforecast.2013.07.001>. URL <http://www.sciencedirect.com/science/article/pii/S0169207013000745>.
- Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, and Rob J. Hyndman. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, 32(3):896 – 913, 2016. ISSN 0169-2070. doi: <https://doi.org/10.1016/j.ijforecast.2016.02.001>. URL <http://www.sciencedirect.com/science/article/pii/S0169207016000133>.
- LV Kantorovich and G Rubinshtein. On a space of totally additive functions, *vestn. Vestn Lening. Univ*, 13:52–59, 1958.
- Daniel Kuhn, Peyman Mohajerin Esfahani, Viet Anh Nguyen, and Soroosh Shafieezadeh-Abadeh. *Wasserstein Distributionally Robust Optimization: Theory and Applications in Machine Learning*, chapter 6, pages 130–166. INFORMS TutORials in Operations Research, 2019. doi: 10.1287/educ.2019.0198. URL <https://pubsonline.informs.org/doi/abs/10.1287/educ.2019.0198>.
- Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1): 115–166, Sep 2018. ISSN 1436-4646. doi: 10.1007/s10107-017-1172-1. URL <https://doi.org/10.1007/s10107-017-1172-1>.

- Theodore S Motzkin, Howard Raiffa, Gerald L Thompson, and Robert M Thrall. The double description method. *Contributions to the Theory of Games*, 2(28):51–73, 1953.
- U Aytun Ozturk, Mainak Mazumdar, and Bryan A Norman. A solution to the stochastic unit commitment problem using chance constrained programming. *IEEE Transactions on Power Systems*, 19(3):1589–1598, 2004.
- Anthony Papavasiliou and Shmuel S Oren. Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network. *Operations Research*, 61(3):578–592, 2013.
- Svetlozar T Rachev and Ludger Rüschemdorf. *Mass Transportation Problems: Volume I: Theory*, volume 1. Springer Science & Business Media, 1998.
- Hamed Rahimian and Sanjay Mehrotra. Distributionally robust optimization: A review. arXiv preprint arXiv:1908.05659, 2019.
- A. Street, F. Oliveira, and J. M. Arroyo. Contingency-constrained unit commitment with  $n - k$  security criterion: A robust optimization approach. *IEEE Transactions on Power Systems*, 26(3):1581–1590, Aug 2011. ISSN 0885-8950. doi: 10.1109/TPWRS.2010.2087367.
- Samer Takriti, John R Birge, and Erik Long. A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems*, 11(3):1497–1508, 1996.
- Qianfan Wang, Jianhui Wang, and Yongpei Guan. Stochastic unit commitment with uncertain demand response. *IEEE Transactions on Power Systems*, 28(1):562–563, 2013.
- Wolfram Wiesemann, Daniel Kuhn, and Melvyn Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014. doi: 10.1287/opre.2014.1314. URL <https://doi.org/10.1287/opre.2014.1314>.
- David Wozabal. A framework for optimization under ambiguity. *Annals of Operations Research*, 193(1):21 – 47, 2012. ISSN 1572-9338. doi: <https://doi.org/10.1007/s10479-010-0812-0>. URL <https://doi.org/10.1007/s10479-010-0812-0>.
- Bo Zeng and Long Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457 – 461, 2013.
- Chaoyue Zhao. *Data-driven risk-averse stochastic program and renewable energy integration*. PhD thesis, University of Florida, 2014.
- Chaoyue Zhao and Yongpei Guan. Data-driven risk-averse stochastic optimization with wasserstein metric. *Operations Research Letters*, 46(2):262 – 267, 2018. ISSN 0167-6377. doi: <https://doi.org/10.1016/j.orl.2018.01.011>. URL <http://www.sciencedirect.com/science/article/pii/S0167637718300506>.
- Chaoyue Zhao, Qianfan Wang, Jianhui Wang, and Yongpei Guan. Expected value and chance constrained stochastic unit commitment ensuring wind power utilization. *IEEE Transactions on Power Systems*, 29(6):2696–2705, 2014.
- A. Zhou, M. Yang, M. Wang, and Y. Zhang. A linear programming approximation of distributionally robust chance-constrained dispatch with wasserstein distance. *IEEE Transactions on Power Systems*, 35(5):3366–3377, 2020.
- R. Zhu, H. Wei, and X. Bai. Wasserstein metric based distributionally robust approximate framework for unit commitment. *IEEE Transactions on Power Systems*, 34(4):2991–3001, 2019.