Tractable Reformulations of Two-Stage Distributionally Robust Linear Programs over the Type-∞ Wasserstein Ball

Weijun Xie*

Department of Industrial & Systems Engineering, Virginia Tech, Blacksburg, VA 24061

Abstract

This paper studies a two-stage distributionally robust stochastic linear program under the type-∞ Wasserstein ball by providing sufficient conditions under which the program can be efficiently computed via a tractable convex program. By exploring the properties of binary variables, the developed reformulation techniques are extended to those with mixed binary random parameters. The main tractable reformulations are projected into the original decision space. The complexity analysis demonstrates that these tractable results are tight under the setting of this paper.

Keywords: Distributionally Robust, Two-stage, Stochastic Program, Tractable, Reformulation.

1. Introduction

1.1. Setting

Consider the two-stage distributionally robust stochastic linear program of the form [25]:

$$v^* = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \mathcal{Z}(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}, Z(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[Z(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \right\}. \tag{1}$$

Above, set $\mathcal{X} \subseteq \mathbb{R}^{n_1}$ denotes the feasible region of the here-and-now decisions $\boldsymbol{x}$, the vector $\boldsymbol{c} \in \mathbb{R}^{n_1}$ denotes the here-and-now objective coefficients, and the function $\mathcal{Z}(\boldsymbol{x})$ denotes the worst-case expected piecewise convex wait-and-see cost function $Z(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})$ (also known as, recourse function) specified by random parameters $\tilde{\boldsymbol{\xi}} \in \Xi$, where its probability distribution $\mathbb{P}$ comes from a family of distributions, denoted by ambiguity set $\mathcal{P}$.

Following the notation in [1, 6, 49], given a realization $\boldsymbol{\xi}$ of $\tilde{\boldsymbol{\xi}}$, we consider the following recourse function:

$$Z(\boldsymbol{x}, \boldsymbol{\xi}) = \min_{\boldsymbol{y}} \left\{ (\boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q})^\top \boldsymbol{y} : \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{x}), \boldsymbol{y} \in \mathbb{R}^{n_2} \right\}. \tag{2}$$

where $\boldsymbol{y}$ represents the second-stage wait-and-see decisions, $\boldsymbol{\xi} = (\boldsymbol{\xi}_q, \boldsymbol{\xi}_T) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$, $\boldsymbol{q} \in \mathbb{R}^{n_2}$ and there are two affine mappings- right-hand mapping $\boldsymbol{h} : \mathbb{R}^{n_1} \to \mathbb{R}^{\ell}$ and technology mapping $\boldsymbol{T} : \mathbb{R}^{n_1} \to \mathbb{R}^{\ell \times m_2}$. Similar to many two-stage stochastic program [6, 49], throughout this paper, we assume that

- (Fixed Recourse) The recourse matrix $\boldsymbol{W} \in \mathbb{R}^{\ell \times n_2}$ is *fixed*; and
- (Separable Uncertainty) The support $\Xi = \Xi_q \times \Xi_T$, where $\Xi_q \subseteq \mathbb{R}^{m_1}, \Xi_T \subseteq \mathbb{R}^{m_2}$.

Both assumptions are quite standard and have appeared in many stochastic programming applications, for example, power systems [16, 23], logistics and supply chain [32, 36], inventory and production [29, 59], agriculture [35], and many others.

The following example illustrates problem (1).

*Corresponding author
Email address: `wxie@vt.edu` (Weijun Xie)

Example 1. (Reliable Facility Location Problem (RFLP) under Probabilistic Disruptions) Let us consider a two-stage facility location problem with random demands and probabilistic disruptions, an extension of the work [15, 36]. Suppose a warehousing company needs to build facilities at candidate locations indexed by $[n_1]$, which are required to serve customers at locations indexed by $[\ell]$. Each facility $s \in [n_1]$ bears a setup cost c_s and due to catastrophic events (e.g., hurricane, power outage, etc.), it might be disrupted, thus, we use $\tilde{\delta}_s \in \{0,1\}$ to denote its status, i.e., $\tilde{\delta}_s = 1$ if it will function well, 0, otherwise. We suppose that each customer $t \in [\ell]$ has a stochastic demand $\tilde{d}_t$ and incurs a unit transportation cost for a shipment from facility $s \in [n_1]$, denoted by $\widehat{c}_{ts}$. The random parameters $\tilde{\boldsymbol{\xi}} = (\tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{d}})$, where its joint probability distribution is usually difficult to characterize. Suppose there are N empirical data points available, denoted by $\{\boldsymbol{\zeta}^j := (\widehat{\boldsymbol{\delta}}^j, \widehat{\boldsymbol{d}}^j)\}_{j\in[N]}$.

To ensure the feasibility of the model, similar to [15, 36], we assume that there is an emergency (or dummy) facility indexed by n_1+1, which will be never disrupted, and its unit transportation cost for each customer $t \in [\ell]$ is $\widehat{c}_{t(n_1+1)} = M$, where M is a large number. Under this setting, distributionally robust RFLP (DR-RFLP) can be formulated as

$$v^* = \min_{\boldsymbol{x}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \mathcal{Z}(\boldsymbol{x}) : \boldsymbol{x} \in \{0,1\}^{n_1}, \mathcal{Z}(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[Z(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})] \right\}, \tag{3a}$$

where the recourse function is

$$Z(\boldsymbol{x}, \boldsymbol{\xi}) = \min_{\boldsymbol{y}} \left\{ \sum_{t\in[\ell]} \sum_{s\in[n_1+1]} \widehat{c}_{ts} d_t y_{ts} : \sum_{s\in[n_1+1]} y_{ts} = 1, y_{ts} \le \delta_s x_s, \forall t \in [\ell], \forall s \in [n_1], \boldsymbol{y} \in \mathbb{R}_+^{\ell\times n_1} \right\}. \tag{3b}$$

1.2. Ambiguity Set

In this paper, we consider $\infty-$Wasserstein ambiguity set $\mathcal{P}$, which is defined as

$$\mathcal{P} = \left\{ \mathbb{P} : \mathbb{P}\left\{ \tilde{\boldsymbol{\xi}} \in \Xi \right\} = 1, W^\infty \left(\mathbb{P}, \mathbb{P}_{\tilde{\boldsymbol{\zeta}}} \right) \le \theta \right\}, \tag{4}$$

where $\infty-$Wasserstein distance [7, 21] is defined as

$$W^\infty (\mathbb{P}_1, \mathbb{P}_2) = \inf_{\mathbb{Q}} \left\{ \text{ess.sup} \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|_p \mathbb{Q}(d\boldsymbol{\xi}_1, d\boldsymbol{\xi}_2) : \begin{array}{l} \mathbb{Q} \text{ is a joint distribution of } \tilde{\boldsymbol{\xi}}_1 \text{ and } \tilde{\boldsymbol{\xi}}_2 \\ \text{with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2 \text{, respectively} \end{array} \right\},$$

ess.sup$(\cdot)$ denotes essential supremum (see [45]), norm $\|\cdot\|_p$ denotes reference distance with $p \in [1,\infty]$ and $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}$ denotes a discrete empirical distribution of $\tilde{\boldsymbol{\zeta}}$ generated by i.i.d. samples $\mathcal{Z} = \{\boldsymbol{\zeta}^j := (\zeta_q^j, \zeta_T^j)\}_{j\in[N]} \subseteq \Xi$ from the true distribution $\mathbb{P}^\infty$, i.e., its point mass function is $\mathbb{P}_{\tilde{\boldsymbol{\zeta}}}\left\{ \tilde{\boldsymbol{\zeta}} = \boldsymbol{\zeta}^j \right\} = \frac{1}{N}$, $\theta \ge 0$ denotes the Wasserstein radius, and $p \ge 1$. Many recent works also studied $\tau-$Wasserstein ambiguity set with $\tau \in [1,\infty)$, where in (4), we replace the $\infty-$Wasserstein distance by the following $\tau-$Wasserstein distance

$$W^\tau (\mathbb{P}_1, \mathbb{P}_2) = \inf_{\mathbb{Q}} \left\{ \sqrt[\tau]{\int_{\Xi\times\Xi} \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|_p^\tau \mathbb{Q}(d\boldsymbol{\xi}_1, d\boldsymbol{\xi}_2)} : \begin{array}{l} \mathbb{Q} \text{ is a joint distribution of } \tilde{\boldsymbol{\xi}}_1 \text{ and } \tilde{\boldsymbol{\xi}}_2 \\ \text{with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2 \text{, respectively} \end{array} \right\}.$$

Clearly, according to [21], $\tau-$Wasserstein distance converges to $\infty-$Wasserstein distance as $\tau \to \infty$. Different types of Wasserstein ambiguity set might provide different tractable results. The results of this paper reveal that $\infty-$Wasserstein ambiguity set indeed delivers more tractable results for problem (1) and it still exhibits attractive convergent properties.

The discussions on advantages of Wasserstein ambiguity sets can be found in [39, 20, 7], which are briefly summarized below: (i) **Data-Driven**. When the number of observed empirical data points grows, the Wasserstein radius shrinks under mild conditions, and thus, the corresponding problem (1) eventually converges to the true two-stage stochastic programming as $N \to \infty$; (ii) **Finite**. It has been shown in [39, 20, 11] that as long as the number of empirical data points is finite, the worst-case probability distribution of the corresponding problem (1) is also finitely supported; and (iii) **Tractability**. There have been many successful developments on tractable reformulations of distributionally robust optimization with Wasserstein ambiguity set, see, for example, [39, 20, 11, 10, 19, 14]. However, for problem (1), the tractable results are quite limited. It is shown that the rate of convergence in τ-Wasserstein distance with $\tau \in [1, \infty)$ of the empirical distribution [18] is similar to that of ∞-Wasserstein distance [50], where the latter requires more restrictive assumption of compact support. On the other hand, if we focus on the convergence of the optimal value v^* to the true optimal one, then the rate of convergence of different Wasserstein distances are similar (see, e.g., [58]). A recent work [58] shows that distributionally robust optimization under ∞-Wasserstein ambiguity set exhibits computational advantages over that under τ-Wasserstein ambiguity set with $\tau \in [1, \infty)$, while by choosing proper Wasserstein radii, different types of ambiguity set are of similar conservatism. Therefore, this paper focuses on developing tractable representations of problem (1) under $\infty-$Wasserstein ambiguity set $\mathcal{P}$, in particular, it focuses on the tractable representations of the worst-case expected wait-and-see cost (i.e., the function $\mathcal{Z}(\boldsymbol{x})$).

1.3. Related Literature

Distributionally robust optimization (DRO) has been used as an alternative modeling paradigm for optimization under uncertainty, where the probability distributions of random parameters are not fully known. Interested readers are referred to [43] for a complete literature review of DRO. Recently, there are several interesting works on exact tractable reformulations of the function $\mathcal{Z}(\boldsymbol{x})$ under three types of ambiguity sets, namely, under moment ambiguity set, phi-divergence based ambiguity set, and Wasserstein ambiguity set.

(i) Moment ambiguity set is specified by the acquired knowledge of some moments (e.g., known first two moments), and has been successfully applied to many different settings (see, for example, [17, 6, 22, 9, 52, 26, 27, 40, 34, 55, 56, 60, 42]). In [17], the authors showed that if the first two moments are known or bounded from above, and the recourse function can be expressed as piecewise maximum of a finite number of functions which are convex in $\boldsymbol{x}$ and concave in the random parameters $\tilde{\boldsymbol{\xi}}$, then the function $\mathcal{Z}(\boldsymbol{x})$ have a tractable representation. In [6], the authors showed that if first two moments are known, then the function $\mathcal{Z}(\boldsymbol{x})$ with only objective uncertainty (i.e., $\tilde{\boldsymbol{\xi}}_T$ is deterministic) can be formulated as a tractable semidefinite program (SDP). The work [40] further showed that if first two moments are known, then the function $\mathcal{Z}(\boldsymbol{x})$ with objective uncertainty and any known support can be reformulated as an SDP, where the positive semidefinite matrix comes from a convex hull of rank-one matrices, and, although computationally intractable in general, the authors were able to establish sufficient conditions under which this SDP formulation becomes tractable.

(ii) Phi-divergence based ambiguity set is specified by the bounded distance between a nominal distribution and true distribution via phi-divergence [2, 3, 28, 30, 31]. In particular, the work [31] showed that for problem (1) with the phi-divergence based ambiguity set can be equivalently reformulated as a convex combination of conditional-value-at-risk and worst-case risk cost, where the tractability follows when both risk measures are tractable.

(iii) Wasserstein ambiguity set is specified by the bounded distance between a nominal distribution and true distribution via Wasserstein metric [39, 11, 10, 12, 13, 14, 20, 19, 25, 7, 37, 54, 57, 61]. In [39] , the authors showed that for problem (1) under $1-$Wasserstein ambiguity set, if the recourse function can be expressed as piecewise maximum of a finite number of functions which are bi-affine in the decision variables $\boldsymbol{x}$ and the random parameters $\tilde{\boldsymbol{\xi}}$, then the function $\mathcal{Z}(\boldsymbol{x})$ has a tractable representation. In [25], the authors extended the tractable results into problem (1) with constraint uncertainty (i.e., $\tilde{\boldsymbol{\xi}}_q$ is deterministic) and $1-$Wasserstein ambiguity set, where the reference distance $\|\cdot\|_1$ and support $\Xi_T = \mathbb{R}^{m_2}$, and proved that for the general problem (1) under Wasserstein ambiguity set, it is in general NP-hard to evaluate the function $\mathcal{Z}(\boldsymbol{x})$. Thereby, the authors proposed a hierarchy of SDP representations to approximate the function $\mathcal{Z}(\boldsymbol{x})$ under $2-$Wasserstein ambiguity set.

Different from [25], this paper focuses on $\infty-$Wasserstein ambiguity set, providing sufficient conditions under which the function $\mathcal{Z}(\boldsymbol{x})$ can be tractable, even with both objective and constraint uncertainties, and further extending the tractable results to the cases where random parameters are mixed binary. As far as the author is concerned, only two works studied $\infty-$Wasserstein ambiguity set, i.e., [7, 8]. The work [7] provided fundamental convergence analysis of $\infty-$Wasserstein ambiguity set, and studied adaptive approximation schemes for the data-driven multi-stage linear program, while the work [8] studied robust two-stage sampling problem with constraint uncertainty and proved that under certain conditions, the proposed multi-policy approximation scheme is asymptotically optimal. Different from these two works, this paper studies problem (1) by exploring exact tractable reformulations of the function $\mathcal{Z}(\boldsymbol{x})$ with $\infty-$Wasserstein ambiguity set and providing the complexity analysis to demonstrate the sharpness of the tractable results, i.e., these tractable results are tight under the setting of this paper.

1.4. Contributions

This paper studies exact reformulations of the worst-case expected wait-and-see cost (i.e., function $\mathcal{Z}(\boldsymbol{x})$) in problem (1) under $\infty-$Wasserstein ambiguity set. The main contributions are highlighted as below.

(i) When random parameters $(\tilde{\boldsymbol{\xi}}_q, \tilde{\boldsymbol{\xi}}_T)$ are continuous, we derive exact tractable reformulations for the function $\mathcal{Z}(\boldsymbol{x})$ with uncertainties in both objective function and constraint system, with objective uncertainty only, as well as with constraint uncertainty only. We prove that our tractable results are sharp.

(ii) When either of random parameters $(\tilde{\boldsymbol{\xi}}_q, \tilde{\boldsymbol{\xi}}_T)$ are binary, by exploring the binary variables in the reformulation, we are able to derive exact tractable reformulations for the function $\mathcal{Z}(\boldsymbol{x})$ under sufficient conditions. Our complexity results show that the tractable results are sharp.

(iii) The main tractable reformulations in this paper are projected to the original decision space, and thus have straightforward interpretations of robustness.

(iv) We demonstrate that if the conditions provided in the above results do not hold, then the proposed reformulations become tractable upper bound and will become exact if the Wasserstein radius goes to zero, i.e., they are asymptotically optimal.

Notation: The following notation is used throughout the paper. We use bold-letters (e.g., $\boldsymbol{x}, \boldsymbol{A}$) to denote vectors or matrices, and use corresponding non-bold letters to denote their components. We let $\boldsymbol{e}$ be the all-one vector or matrix whenever necessary, let $\mathbf{0}$ be the all-zero vector or matrix whenever necessary, and we let $\boldsymbol{e}_i$ be the ith standard basis vector. Given an integer n, we let $[n] := \{1, 2, \ldots, n\}$, and use $\mathbb{R}^n_+ := \{\boldsymbol{x} \in \mathbb{R}^n : x_l \geq 0, \forall l \in [n]\}$ and $\mathbb{R}^n_- := \{\boldsymbol{x} \in \mathbb{R}^n : x_l \leq 0, \forall l \in [n]\}$. Given a real number t, we let $(t)_+ := \max\{t, 0\}$. Given a finite set I, we let $|I|$ denote its cardinality. We let $\tilde{\boldsymbol{\xi}}$ denote a random vector with support Ξ and denote one of its realization by $\boldsymbol{\xi}$. Given a real-valued random variable $\tilde{\xi} : \Omega \to \mathbb{R}$ with probability distribution $\mathbb{P}$, its ess.sup$(X) := \inf\{c : \mathbb{P}\{\omega : \tilde{\xi}(\omega) > c\} = 1\}$. Given a set R, the characteristic function $\chi_R(\boldsymbol{x}) = 0$ if $\boldsymbol{x} \in R$, and ∞, otherwise, while the indicator function $\mathbb{I}(\boldsymbol{x} \in R)$ =1 if $\boldsymbol{x} \in R$, and 0, otherwise. We let $\boldsymbol{I}_n$ denote $n \times n$ identify matrix. For a vector $\boldsymbol{a}$, we let $|\boldsymbol{a}|$ denote the result by taking element-wise absolute and let $(\boldsymbol{a})_+ = \max\{\boldsymbol{a}, 0\}$ by taking element-wise maximum. For a matrix $\boldsymbol{A}$, we let $|\boldsymbol{A}|$ denote the result by taking element-wise absolute, let $(\boldsymbol{A})_+ = \max\{\boldsymbol{A}, 0\}$ by taking element-wise maximum, and let $\|\boldsymbol{A}\|_p$ denote its element-wise p-norm with $p \in [1, \infty]$. Additional notation will be introduced as needed.

2. Preliminaries

Similar to [25], we will make the following assumption throughout this paper.

- (Sufficiently Expensive Recourse) For any $\boldsymbol{x} \in \mathcal{X}$, the dual of the second-stage problem (2) is feasible for all $\boldsymbol{\xi} \in \Xi$.

Note that this assumption is used to ensure that the strong duality of the second-stage problem (2) always holds. If this assumption does not hold, then the proposed reformulations in this paper might not be exact.

It has been shown in [7] that

Lemma 1. *(Proposition 3 in [7]) the $\infty-$Wasserstein ambiguity set $\mathcal{P}_\infty$ has the following equivalent form*

$$\mathcal{P} = \left\{ \frac{1}{N} \sum_{k\in[N]} \Delta((\tilde{\boldsymbol{\xi}}_q, \tilde{\boldsymbol{\xi}}_T) - (\boldsymbol{\xi}_q^k, \boldsymbol{\xi}_T^k)) : \exists(\boldsymbol{\xi}_q^k, \boldsymbol{\xi}_T^k), \|(\boldsymbol{\xi}_q^k, \boldsymbol{\xi}_T^k) - (\boldsymbol{\zeta}_q^k, \boldsymbol{\zeta}_T^k)\|_p \le \theta, \forall k \in [N] \right\}, \tag{5}$$

where $\Delta(\cdot)$ is the Dirac delta function.

This representation has the following interpretation, i.e., worst-case distributions are also supported by N points and each point can only deviate θ from one of the empirical data $\{(\boldsymbol{\zeta}_q^k, \boldsymbol{\zeta}_T^k)\}_{k\in[N]}$.

According to the strong duality of distributionally robust optimization with $\infty-$Wasserstein ambiguity set [7], we observe that the function $\mathcal{Z}(\boldsymbol{x})$ can be equivalently represented as the following bilinear program.

Lemma 2. *The function $\mathcal{Z}(\boldsymbol{x})$ is equivalent to*

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j\in[N]} \sup_{(\boldsymbol{\xi}_q, \boldsymbol{\xi}_T)\in\Xi, \boldsymbol{\pi}\in\mathbb{R}_+^\ell} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T)^\top \boldsymbol{\pi} : \|(\boldsymbol{\xi}_q, \boldsymbol{\xi}_T) - (\boldsymbol{\zeta}_q^j, \boldsymbol{\zeta}_T^j)\|_p \le \theta, \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \right\}. \tag{6}$$

Proof: According to equivalent representation (5) of $\mathcal{P}$, $\mathcal{Z}(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[Z(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})]$ is equivalent to

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j\in[N]} \sup_{\boldsymbol{\xi}} \left\{ Z(\boldsymbol{x}, \boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi, \|\boldsymbol{\xi} - \boldsymbol{\zeta}^j\|_p \le \theta \right\}. \tag{7a}$$

Suppose $\boldsymbol{\pi}$ is the dual vector associated with constraints (2) then we can equivalently represent $Z(\boldsymbol{x}, \boldsymbol{\xi})$ as

$$Z(\boldsymbol{x}, \boldsymbol{\xi}) = \max_{\boldsymbol{\pi}} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T)^\top \boldsymbol{\pi} : \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q}, \boldsymbol{\pi} \in \mathbb{R}_+^\ell \right\}. \tag{7b}$$

Substituting (7b) into (7a) and using the fact that $\boldsymbol{\xi} = (\boldsymbol{\xi}_q, \boldsymbol{\xi}_T)$ and $\boldsymbol{\zeta}^j = (\boldsymbol{\zeta}_q^j, \boldsymbol{\zeta}_T^j)$, we arrive at (6).

□

Note that (i) the inner supremum of (6) is to maximize bilinear objective function over convex constraints, which is often difficult to solve. Therefore, the main focus of this paper is to study the complexity of evaluating the function $\mathcal{Z}(\boldsymbol{x})$ and provide sufficient conditions under which the inner supremum is efficiently solvable; and (ii) The proof of Lemma 2 inspires that if $Z(\boldsymbol{x}, \boldsymbol{\xi})$ is Lipschitz continuous, then v^* is lower and upper bounded by its sampling average approximation (SAA) counterparts. Thus, the rate of convergence results for SAA [49, 24] simply apply to problem (1) if we choose $\theta = O(N^{-1/2})$. This observation is summarized below.

Proposition 1. *Suppose that $|Z(\boldsymbol{x}, \boldsymbol{\xi}) - Z(\boldsymbol{x}, \boldsymbol{\zeta})| \le L\|\boldsymbol{\xi} - \boldsymbol{\zeta}\|_p$ for all $\boldsymbol{x} \in X, \boldsymbol{\xi}, \boldsymbol{\zeta} \in \Xi$. Then, we have*

$$v^{\text{SAA}} \le v^* \le v^{\text{SAA}} + L\theta$$

where $v^{\text{SAA}} := \min_{\boldsymbol{x}\in X} \boldsymbol{c}^\top \boldsymbol{x} + \sum_{k\in[N]} Z(\boldsymbol{x}, \boldsymbol{\zeta}^k)$. Thus, the rate of convergence results for SAA [49, 24] simply apply to problem (1) *if we choose $\theta = O(N^{-1/2})$.*

Proof: As empirical distribution $\mathbb{P}_{\tilde{\zeta}}$ is feasible to $\mathcal{P}$, thus in problem (1), we must have $v^{\mathrm{SAA}} \leq v^*$.

According to (7a) and the assumption that $|Z(\boldsymbol{x},\boldsymbol{\xi}) - Z(\boldsymbol{x},\boldsymbol{\zeta})| \leq L\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|_p$ for all $\boldsymbol{x} \in X, \boldsymbol{\xi},\boldsymbol{\zeta} \in \Xi$, we have

$$\mathcal{Z}(\boldsymbol{x}) \leq \frac{1}{N}\sum_{j\in[N]} Z(\boldsymbol{x},\boldsymbol{\zeta}^j) + \frac{1}{N}\sum_{j\in[N]} \sup_{\boldsymbol{\xi}\in\Xi} \left\{ L\|\boldsymbol{\xi}-\boldsymbol{\zeta}^j\|_p : \|\boldsymbol{\xi}-\boldsymbol{\zeta}^j\|_p \leq \theta \right\} \leq v^{\mathrm{SAA}} + L\theta,$$

where the second inequality is due to $\|\boldsymbol{\xi}-\boldsymbol{\zeta}^j\|_p \leq \theta$. □

Other useful tools that this paper relies on are summarized below.

Property 1. *(i) (Dual Norm, [44]) For any norm $\|\cdot\|_p$ with $p \in [1,\infty]$, its dual norm is*

$$\|\boldsymbol{r}\|_{p*} = \max_{\boldsymbol{s}} \left\{ \boldsymbol{r}^\top \boldsymbol{s} : \|\boldsymbol{s}\|_p \leq 1 \right\},$$

where $p* = \frac{p}{p-1}$;

(ii) (Integral Polyhedron, [46]) Given a rational polyhedron $P = \{\boldsymbol{r} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{r} \geq \boldsymbol{b}\}$ is integral if and only if $P = \mathrm{conv}(P \cap \mathbb{Z}^n)$;

(iii) (Tractability, [4]) We say the function $\mathcal{Z}(\boldsymbol{x})$ has a tractable representation, if for any given $\boldsymbol{x} \in \mathbb{R}^{n_1}$, there exists an efficient algorithm which can evaluate the function $\mathcal{Z}(\boldsymbol{x})$ in time polynomial in $n_1, n_2, m_2, m_2, \ell, N$.

3. Continuous Support: Tractable Reformulations and Complexity Analysis

In this section, we consider the random parameters to be continuous, i.e., both $\tilde{\boldsymbol{\xi}}_q, \tilde{\boldsymbol{\xi}}_T$ are continuous. For example, in the newsvendor problem studied by [41], the authors considered random supply and demand, where both are continuous. We first provide the tractable representations of the function $\mathcal{Z}(\boldsymbol{x})$ under various settings and then show that in general, it is NP-hard to evaluate the function $\mathcal{Z}(\boldsymbol{x})$. We split this section into four parts, which include tractable reformulations of general problem (1), special problem (1) with objective uncertainty only, special problem (1) with constraint uncertainty only, and complexity analysis.

3.1. Tractable Reformulation I: General Problem (1) with L_∞ Reference Distance

For the general problem (1), we show that the function $\mathcal{Z}(\boldsymbol{x})$ has a tractable representation given that the reference distance is $\|\cdot\|_p = \|\cdot\|_\infty$ (i.e., $p=\infty$) and the image of the technology mapping $\boldsymbol{T}(\boldsymbol{x})$ is always non-negative or non-positive.

Theorem 1. *Suppose that $\Xi = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$. If $p = \infty$ and $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_+^{\ell\times m_1}$ or $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_-^{\ell\times m_1}$, then the function $\mathcal{Z}(\boldsymbol{x})$ is equivalent to*

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \min_{\boldsymbol{y}\in\mathbb{R}^{n_2}} \left\{ (\boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q})^\top \boldsymbol{y} + \theta\|\boldsymbol{Q}^\top\boldsymbol{y}\|_1 : \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j + \boldsymbol{W}\boldsymbol{y} - \theta|\boldsymbol{T}(\boldsymbol{x})|\boldsymbol{e} \geq \boldsymbol{h}(\boldsymbol{x}) \right\}. \tag{8}$$

Proof: Since $\Xi = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$ and $p = \infty$, thus (6) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \sup_{\boldsymbol{\pi}\in\mathbb{R}_+^\ell,\boldsymbol{\xi}_q,\boldsymbol{\xi}_T} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T)^\top \boldsymbol{\pi} : \|\boldsymbol{\xi}_q - \boldsymbol{\zeta}_q^j\|_\infty \leq \theta, \|\boldsymbol{\xi}_T - \boldsymbol{\zeta}_T^j\|_\infty \leq \theta, \boldsymbol{W}^\top\boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \right\}.$$

Above, optimizing $\boldsymbol{\xi}_T$ and using the dual norm of $\|\cdot\|_\infty$, we have

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \sup_{\boldsymbol{\pi}\in\mathbb{R}_+^\ell,\boldsymbol{\xi}_q} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j)^\top \boldsymbol{\pi} + \theta\|\boldsymbol{T}(\boldsymbol{x})^\top\boldsymbol{\pi}\|_1 : \|\boldsymbol{\xi}_q - \boldsymbol{\zeta}_q^j\|_\infty \leq \theta, \boldsymbol{W}^\top\boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \right\}. \tag{9}$$

Note that since $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_+^{\ell \times m_1}$ or $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_-^{\ell \times m_1}$, thus $\|\boldsymbol{T}(\boldsymbol{x})^\top \boldsymbol{\pi}\|_1 = \boldsymbol{e}^\top |\boldsymbol{T}(\boldsymbol{x})|^\top \boldsymbol{\pi}$. Let $\boldsymbol{y}$ denote the dual variables of the constraints $\boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q}$. Then according to the strong duality of linear programming, (9) is equivalent to

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j \in [N]} \min_{\boldsymbol{y} \in \mathbb{R}^{n_2}} \sup_{\boldsymbol{\pi} \in \mathbb{R}_+^{\ell}, \boldsymbol{\xi}_q} \Big\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j)^\top \boldsymbol{\pi} + \theta \boldsymbol{e}^\top |\boldsymbol{T}(\boldsymbol{x})|^\top \boldsymbol{\pi} + \boldsymbol{y}^\top (\boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} - \boldsymbol{W}^\top \boldsymbol{\pi}) : \\ \|\boldsymbol{\xi}_q - \boldsymbol{\zeta}_q^j\|_\infty \le \theta \Big\},$$

which is equivalent to (8) by optimizing over $(\boldsymbol{\xi}_q, \boldsymbol{\pi})$. □

We make the following remarks about Theorem 1 and its corresponding formulation (8).

(i) We can introduce auxiliary variables to linearize the terms $\|\boldsymbol{Q}^\top \boldsymbol{y}\|_1, |\boldsymbol{T}(\boldsymbol{x})|$ and reformulate the minimization problem (8) as a linear program;

(ii) If $\theta = 0$, i.e., if the empirical distribution is sufficient to describe the probability of random parameters, then $\mathcal{Z}(\boldsymbol{x}) = 1/N \sum_{j \in [N]} Z(\boldsymbol{x}, \boldsymbol{\zeta}^j)$;

(iii) The extra terms, $\theta \|\boldsymbol{Q}^\top \boldsymbol{y}\|_1$ in the objective and $-\theta |\boldsymbol{T}(\boldsymbol{x})|_1 \boldsymbol{e}$ in the constraints, enforce the robustness of the proposed formulation due to ambiguous distributional information. These terms will vanish if more and more observations have been made to drive the Wasserstein radius to be 0. The extra term $\theta \|\boldsymbol{Q}^\top \boldsymbol{y}\|_1$ in the objective can be also interpreted as a "regularizer", which has been discovered in [19, 10] for DRO under τ- Wasserstein balls. For more discussions about asymptotic behavior of Wasserstein ambiguity sets, interested readers are referred to [8, 7, 11, 39, 25, 54];

(iv) If the assumption that $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_+^{\ell \times m_1}$ or $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_-^{\ell \times m_1}$ does not hold, then (8) provides an upper bound for $\mathcal{Z}(\boldsymbol{x})$ due to the triangle inequality $\theta \|\boldsymbol{T}(\boldsymbol{x})^\top \boldsymbol{\pi}\|_1 \le \theta \boldsymbol{e}^\top |\boldsymbol{T}(\boldsymbol{x})|^\top \boldsymbol{\pi}$ and this upper bound will become exact when $\theta \to 0$ since this approximation term is associated with Wasserstein radius θ and thus will vanish as $\theta \to 0$; and

(v) Similarly, if the reference distance is defined by other norm $\|\cdot\|_p$ such that $p \in [1, \infty)$, then according to the following formula

$$\|\boldsymbol{\xi}\|_p \le \sqrt[p]{m_1 + m_2} \|\boldsymbol{\xi}\|_\infty.$$

Thus, (8) provides an upper bound for $\mathcal{Z}(\boldsymbol{x})$ by inflating θ to $\sqrt[p]{m_1 + m_2}\theta$ and this upper bound will become exact when $\theta \to 0$.

According to the representation result in Theorem 1, we provide the following equivalent deterministic reformulation of problem (1).

Proposition 2. *Suppose that $\Xi = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$. If $p = \infty$ and $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_+^{\ell \times m_1}$ or $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_-^{\ell \times m_1}$, then problem* (1) *is equivalent to*

$$v^* = \min_{\boldsymbol{x}, \boldsymbol{y}} \quad \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{N} \sum_{j \in [N]} [(\boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q})^\top \boldsymbol{y}^j + \theta \|\boldsymbol{Q}^\top \boldsymbol{y}^j\|_1], \tag{10a}$$

$$s.t. \quad \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j + \boldsymbol{W}\boldsymbol{y}^j - \theta |\boldsymbol{T}(\boldsymbol{x})| \boldsymbol{e} \ge \boldsymbol{h}(\boldsymbol{x}), \forall j \in [N], \tag{10b}$$

$$\boldsymbol{x} \in \mathcal{X}, \boldsymbol{y}^j \in \mathbb{R}^{n_2}, \forall j \in [N]. \tag{10c}$$

The following example illustrates how to use the proposed formulation in DR-RFLP.

Example 2. Consider DR-RFLP in Example 1. Suppose the reference distance is $\|\cdot\|_\infty$ and the support of $\tilde{\boldsymbol{\xi}}$ is $\mathbb{R}^{n_1} \times \mathbb{R}^{\ell}$. Since the coefficients of uncertain parameters $\tilde{\boldsymbol{\delta}}$ in the constraints (3b) always have the same sign, according to Proposition 2, DR-RFLP can be equivalently formulated as the following mixed integer linear program (MILP):

$$v^* = \min_{\boldsymbol{x},\boldsymbol{y}} \ \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{N} \sum_{j\in[N]} \sum_{t\in[\ell]} \sum_{s\in[n_1+1]} \widehat{c}_{ts}(\widehat{d}_t^j + \theta) y_{ts}^j, \tag{11a}$$

$$\text{s.t.} \sum_{s\in[n_1+1]} y_{ts}^j = 1, \forall j \in [N], \forall t \in [\ell], \tag{11b}$$

$$y_{ts}^j \leq (\widehat{\delta}_s^j - \theta) x_s, \forall j \in [N], \forall t \in [\ell], \forall s \in [n_1], \tag{11c}$$

$$\boldsymbol{x} \in \{0,1\}^{n_1}, \boldsymbol{y}^j \in \mathbb{R}_+^{\ell \times n_1}, \forall j \in [N]. \tag{11d}$$

3.2. Tractable Reformulation II: With Objective Uncertainty Only

If there are only objective uncertainty involved in problem (1), then the function $\mathcal{Z}(\boldsymbol{x})$ always has a tractable representation provided that the reference distance is $\|\cdot\|_p$ for any $p \in [1, \infty]$.

Theorem 2. *Suppose that $\Xi = \mathbb{R}^{m_1} \times \{\boldsymbol{\xi}_T\}$. Then for any $p \in [1, \infty]$, the function $\mathcal{Z}(\boldsymbol{x})$ is equivalent to*

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j\in[N]} \min_{\boldsymbol{y}\in\mathbb{R}^{n_2}} \left\{ (\boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q})^\top \boldsymbol{y} + \theta \|\boldsymbol{Q}^\top \boldsymbol{y}\|_{p*} : \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{x}) \right\}, \tag{12}$$

where $\|\cdot\|_{p*}$ denotes the dual norm of $\|\cdot\|_p$ with $p* = \frac{p}{p-1}$.

Proof: Since $\Xi = \mathbb{R}^{m_1} \times \{\boldsymbol{\xi}_T\}$, (6) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j\in[N]} \sup_{\boldsymbol{\pi}\in\mathbb{R}_+^\ell, \boldsymbol{\xi}_q} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T)^\top \boldsymbol{\pi} : \|\boldsymbol{\xi}_q - \boldsymbol{\zeta}_q^j\|_p \leq \theta, \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \right\}. \tag{13}$$

Let $\boldsymbol{y}$ denote the dual variables of the constraints $\boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q}$. Since the inner supremum of (13) is essentially strictly feasible, according to the strong duality of conic programming [5], (13) is equivalent to

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j\in[N]} \min_{\boldsymbol{y}\in\mathbb{R}^{n_2}} \sup_{\boldsymbol{\pi}\in\mathbb{R}_+^\ell, \boldsymbol{\xi}_q} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T)^\top \boldsymbol{\pi} + \boldsymbol{y}^\top (\boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} - \boldsymbol{W}^\top \boldsymbol{\pi}) : \|\boldsymbol{\xi}_q - \boldsymbol{\zeta}_q^j\|_p \leq \theta \right\},$$

which is further equivalent to (12) by optimizing over $(\boldsymbol{\pi}, \boldsymbol{\xi}_q)$. □

We make the following remarks about Theorem 2 and its corresponding formulation (12).

(i) For any rational $p \in [1, \infty]$, the penalty term $\theta\|\boldsymbol{Q}^\top \boldsymbol{y}\|_{p*}$ is second order conic representable [5]. Therefore, (12) can be further reformulated as a second order conic program; and

(ii) The penalty term, $\theta\|\boldsymbol{Q}^\top \boldsymbol{y}\|_{p*}$ in the objective, enforces the robustness of the proposed model due to ambiguous distributional information, which has been observed by [19, 10, 47, 48] for DRO under τ-Wasserstein balls. This term will vanish if more and more observations have been made to drive the Wasserstein radius to 0.

We provide the following equivalent deterministic reformulation of problem (1) with objective uncertainty only.

Proposition 3. *Suppose that $\Xi = \mathbb{R}^{m_1} \times \{\boldsymbol{\xi}_T\}$. Then for any $p \in [1, \infty]$, problem* (1) *is equivalent to*

$$v^* = \min_{\boldsymbol{x},\boldsymbol{y}} \quad \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{N} \sum_{j\in[N]} \left[(\boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q})^\top \boldsymbol{y}^j + \theta \|\boldsymbol{Q}^\top \boldsymbol{y}^j\|_{p*} \right], \tag{14a}$$

$$s.t. \quad \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T + \boldsymbol{W}\boldsymbol{y}^j \geq \boldsymbol{h}(\boldsymbol{x}), \forall j \in [N], \tag{14b}$$

$$\boldsymbol{x} \in \mathcal{X}, \boldsymbol{y}^j \in \mathbb{R}^{n_2}, \forall j \in [N]. \tag{14c}$$

3.3. Tractable Reformulation III: With Constraint Uncertainty Only

If there are only constraint uncertainty involved in problem (1), then the function $\mathcal{Z}(\boldsymbol{x})$ can have a tractable representation given that the reference distance when $p = 1$.

Theorem 3. *Suppose that $\Xi = \{\boldsymbol{\xi}_q\} \times \mathbb{R}^{m_2}$ and $p = 1$. Then the function $\mathcal{Z}(\boldsymbol{x})$ is equivalent to*

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j \in [N]} \max_{r \in \{-1,1\}} \max_{i \in [m_2]} \min_{\boldsymbol{y} \in \mathbb{R}^{n_2}} \left\{ (\boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q})^\top \boldsymbol{y} : \boldsymbol{T}(\boldsymbol{x})\zeta_T^j + \boldsymbol{W}\boldsymbol{y} - \theta r \boldsymbol{T}(\boldsymbol{x})\boldsymbol{e}_i \geq \boldsymbol{h}(\boldsymbol{x}) \right\}. \tag{15}$$

Proof: See Appendix A.1. □

We make the following remarks about Theorem 3 and its corresponding formulation (15).

(i) Clearly, since problem (1) with constraint uncertainty only is a special case of general problem (1), thus the result from Theorem 1 directly follows and is not listed here;

(ii) In [25], the authors also proved that under the setting of Theorem 3, problem (1) with 1-Wasserstein ambiguity set is tractable. However, our formulation and required proof technique are quite different from theirs;

(iii) To obtain $\mathcal{Z}(\boldsymbol{x})$, one needs to solve $2m_1$ linear programs for each $j \in [N]$;

(iv) If $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_+^{\ell \times m_1}$ or $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_-^{\ell \times m_1}$, then due to monotonicity, we must have optimal $r^* = 1$ or $r^* = -1$, respectively. Thus, for these cases, one only needs to solve m_1 linear programs instead of $2m_1$ for each $j \in [N]$; and

(v) The penalty term, $-\theta r \boldsymbol{T}(\boldsymbol{x})\boldsymbol{e}_i$ in the constraints, enforces the robustness of the proposed model due to ambiguous distributional information.

In view of the result in Theorem 3, we provide the following equivalent deterministic reformulation of problem (1).

Proposition 4. *Suppose that $\Xi = \{\boldsymbol{\xi}_q\} \times \mathbb{R}^{m_2}$ and $p = 1$. Then problem* (1) *is equivalent to*

$$\begin{aligned} v^* = \min_{\boldsymbol{x}, \boldsymbol{\eta}} \quad & \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{N} \sum_{j \in [N]} \eta_j, \\ \text{s.t.} \quad & \eta_j \geq (\boldsymbol{Q}\boldsymbol{\xi}_q^j + \boldsymbol{q})^\top \boldsymbol{y}^{ijr}, \forall j \in [N], \forall i \in [m_1], \forall r \in \{-1,1\}, \\ & \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j + \boldsymbol{W}\boldsymbol{y}^{ijr} - \theta r \boldsymbol{T}(\boldsymbol{x})\boldsymbol{e}_i \geq \boldsymbol{h}(\boldsymbol{x}), \forall j \in [N], \forall i \in [m_1], \forall r \in \{-1,1\}, \\ & \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y}^{ijr} \in \mathbb{R}^{n_2}, \forall j \in [N], \forall i \in [m_1], \forall r \in \{-1,1\}. \end{aligned}$$

Another special case of problem (1) without objective uncertainty is that the dual constraint system of (2) is bounded and has a small number of extreme points. In this case, equivalently, we can represent the recourse function in the form of piece-wise max of a finite number of affine functions in the random parameters, and obtain the tractable reformulation for any reference distance $\|\cdot\|_p$ for any $p \in [1, \infty]$. This result is summarized below.

Proposition 5. *Suppose that $\Xi = \mathbb{R}^\tau$ and $z(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[\max_{i \in [m]} \{\boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + d_i(\boldsymbol{x})\}]$ with affine functions $\boldsymbol{a}_i(\boldsymbol{x}) : \mathbb{R}^{n_1} \to \mathbb{R}^\tau$ and $d_i(\boldsymbol{x}) : \mathbb{R}^{n_1} \to \mathbb{R}$ for each $i \in [m]$. Then*

- *Function $z(\boldsymbol{x})$ is equivalent to*

$$z(\boldsymbol{x}) = \frac{1}{N} \sum_{j \in [N]} \max_{i \in [m]} \left[\boldsymbol{a}_i(\boldsymbol{x})^\top \widehat{\boldsymbol{\zeta}}^j + d_i + \theta \|\boldsymbol{a}_i(\boldsymbol{x})\|_{p*} \right]. \tag{16}$$

- *Problem* (1) *is equivalent to*

$$v^* = \min_{\boldsymbol{x},\boldsymbol{\eta}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{N}\sum_{j\in[N]} \eta_j : \eta_j \geq \boldsymbol{a}_i(\boldsymbol{x})^\top \widehat{\boldsymbol{\zeta}}^j + d_i + \theta \|\boldsymbol{a}_i(\boldsymbol{x})\|_{p*}, \forall j \in [N], \forall i \in [m], \boldsymbol{x} \in \mathcal{X} \right\}. \quad (17)$$

Proof: Since $\Xi = \mathbb{R}^\tau$ and $Z(\boldsymbol{x}, \boldsymbol{\xi}) = \max_{i\in[m]}\{\boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + d_i(\boldsymbol{x})\}$, (7a) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \max_{i\in[m]} \sup_{\boldsymbol{\xi}} \left\{ \boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + d_i(\boldsymbol{x}) : \|\boldsymbol{\xi} - \boldsymbol{\zeta}^j\|_p \leq \theta \right\},$$

Above, optimizing $\boldsymbol{\xi}$ using dual norm of $\|\cdot\|_p$, we arrive at (16).

The formulation (17) follows from a straightforward linearization. □

3.4. Complexity Analysis

We close this section by showing that for general reference distance $\|\cdot\|_p$ with $p \in (1, \infty]$, computing the function $\mathcal{Z}(\boldsymbol{x})$ with $N = 1$ is NP-hard.

Proposition 6. *Computing $\mathcal{Z}(\boldsymbol{x})$ is NP-hard whenever the reference distance is $\|\cdot\|_p$ with any $p \in (1, \infty]$, $N = 1$, $\Xi = \{\boldsymbol{\xi}_q\} \times \mathbb{R}^{m_2}$, $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{\zeta}_T^1 = \boldsymbol{0}$, and Wasserstein radius $\theta > 0$.*

Proof: See Appendix A.2. □

This complexity result suggests that the tractable results obtained in this section are the best we can get.

4. Binary Support: Tractable Reformulations and Complexity Analysis

In this section, we consider one of the random parameters to be continuous, i.e., either of $\tilde{\boldsymbol{\xi}}_q, \tilde{\boldsymbol{\xi}}_T$ is continuous. Some practical stochastic programming applications might involve binary random parameters. For instance, in the reliable facility location problem with probabilistic disruptions [15, 36], the disruption parameters are in fact binary, i.e., $\mathbb{P}\{\tilde{\boldsymbol{\delta}} \in \{0,1\}^{n_1}\} = 1$; in the stochastic power systems with contingencies [51, 53], the availability of a system component is also binary supported. Motivated by these applications, in this section, we explore the tractable representations of the function $\mathcal{Z}(\boldsymbol{x})$ when one of random parameters $\tilde{\boldsymbol{\xi}}_q, \tilde{\boldsymbol{\xi}}_T$ is binary, i.e., we consider either $\tilde{\boldsymbol{\xi}}_q \in \{0,1\}^{m_1}$ or $\tilde{\boldsymbol{\xi}}_T \in \{0,1\}^{m_2}$, and the other random parameters are continuous. Our complexity analysis shows that it is unlikely to obtain any tractable results when both random parameters $\tilde{\boldsymbol{\xi}}_q, \tilde{\boldsymbol{\xi}}_T$ are binary, and thus we leave it to interested readers. Note that different from the setting of this paper, robust programs with the mixed-integer uncertainty set were studied in a recent work [38].

4.1. Tractable Reformulation I: General Problem (1) *with L_∞ Reference Distance*

For the general problem (1) with objective uncertainty, the function $\mathcal{Z}(\boldsymbol{x})$ has a tractable representation given that the reference distance is $\|\cdot\|_p = \|\cdot\|_\infty$ (i.e., $p = \infty$).

Theorem 4. *Suppose $p = \infty$ and $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_+^{\ell\times m_1}$ or $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_-^{\ell\times m_1}$.*

(i) If $\Xi = \mathbb{R}^{m_1} \times \{0,1\}^{m_2}$, then the function $\mathcal{Z}(\boldsymbol{x})$ is equivalent to

$$\mathcal{Z}(\boldsymbol{x}) = \begin{cases} \frac{1}{N}\sum_{j\in[N]} \min_{\boldsymbol{y}\in\mathbb{R}^{n_2}} \left\{ (\boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q})^\top \boldsymbol{y} + \theta\|\boldsymbol{Q}^\top \boldsymbol{y}\|_1 : -(-\boldsymbol{T}(\boldsymbol{x}))_+ \boldsymbol{e} + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{x}) \right\}, & \text{if } \theta \geq 1 \\ \frac{1}{N}\sum_{j\in[N]} \min_{\boldsymbol{y}\in\mathbb{R}^{n_2}} \left\{ (\boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q})^\top \boldsymbol{y} + \theta\|\boldsymbol{Q}^\top \boldsymbol{y}\|_1 : \boldsymbol{T}(\boldsymbol{x})\zeta_T^j + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{x}) \right\}, & \text{if } \theta < 1 \end{cases}; \quad (18)$$

(ii) If $\Xi = \{0,1\}^{m_1} \times \mathbb{R}^{m_2}$ and the polyhedron $\left\{(\boldsymbol{\pi}, \boldsymbol{\xi}_q) \in \mathbb{R}_+^{\ell} \times [0,1]^{m_1} : \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q}\right\}$ is integral, then the function $\mathcal{Z}(\boldsymbol{x})$ is equivalent to

$$\mathcal{Z}(\boldsymbol{x}) = \begin{cases} \frac{1}{N} \sum_{j \in [N]} \min_{\boldsymbol{y} \in \mathbb{R}^{n_2}} \left\{ (\boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q})^\top \boldsymbol{y} : \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j + \boldsymbol{W}\boldsymbol{y} - \theta |\boldsymbol{T}(\boldsymbol{x})| \boldsymbol{e} \geq \boldsymbol{h}(\boldsymbol{x}) \right\}, & \text{if } \theta \geq 1 \\ \frac{1}{N} \sum_{j \in [N]} \min_{\boldsymbol{y} \in \mathbb{R}^{n_2}} \left\{ (\boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q})^\top \boldsymbol{y} + \boldsymbol{e}^\top (\boldsymbol{Q}^\top \boldsymbol{y})_+ : \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j + \boldsymbol{W}\boldsymbol{y} - \theta |\boldsymbol{T}(\boldsymbol{x})| \boldsymbol{e} \geq \boldsymbol{h}(\boldsymbol{x}) \right\}, & \text{if } \theta < 1 \end{cases}. \tag{19}$$

Proof: See Appendix A.3. □

We make the following remarks about Theorem 4 and its corresponding formulations (18) and (19).

(i) We can introduce auxiliary variables to linearize the terms $\|\boldsymbol{Q}^\top \boldsymbol{y}\|_1, |\boldsymbol{T}(\boldsymbol{x})|, (\boldsymbol{Q}^\top \boldsymbol{y})_+$ and reformulate the minimization problems (18) and (19) as linear programs;

(ii) The function $\mathcal{Z}(\boldsymbol{x})$ depends on whether the Wasserstein radius θ is greater than 1 or not. Particularly, when $\theta < 1$, the radius is too small to generate any new adversarial samples, as the support is restricted to the binary points;

(iii) If the assumption that $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_+^{\ell \times m_1}$ or $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_-^{\ell \times m_1}$ does not hold, then (18) provides an upper bound for $\mathcal{Z}(\boldsymbol{x})$ and this upper bound will become exact when $\theta \to 0$; and

(iv) If one of assumptions that (1) $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_+^{\ell \times m_1}$ or $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_-^{\ell \times m_1}$; and (2) the polyhedron $\{(\boldsymbol{\pi}, \boldsymbol{\xi}_q) \in \mathbb{R}_+^{\ell} \times [0,1]^{m_1} : \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q}\}$ is integral, does not hold, then (19) provides an upper bound for $\mathcal{Z}(\boldsymbol{x})$ and this upper bound will become exact when $\theta \to 0$.

According to the representation results in Theorem 4, we provide the following equivalent deterministic reformulation of problem (1).

Proposition 7. *Suppose $p = \infty$, and $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_+^{\ell \times m_1}$ or $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_-^{\ell \times m_1}$.*

(i) If $\Xi = \mathbb{R}^{m_1} \times \{0,1\}^{m_2}$, then problem (1) is equivalent to

$$v^* = \min_{\boldsymbol{x}, \boldsymbol{y}} \quad \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{N} \sum_{j \in [N]} [(\boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q})^\top \boldsymbol{y}^j + \theta \|\boldsymbol{Q}^\top \boldsymbol{y}^j\|_1], \tag{20a}$$

$$\text{s.t.} \quad \begin{matrix} -(-\boldsymbol{T}(\boldsymbol{x}))_+ \boldsymbol{e} + \boldsymbol{W}\boldsymbol{y}^j \geq \boldsymbol{h}(\boldsymbol{x}), \forall j \in [N], & \text{if } \theta \geq 1 \\ \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j + \boldsymbol{W}\boldsymbol{y}^j \geq \boldsymbol{h}(\boldsymbol{x}), \forall j \in [N], & \text{if } \theta < 1 \end{matrix}, \tag{20b}$$

$$\boldsymbol{x} \in \mathcal{X}, \boldsymbol{y}^j \in \mathbb{R}^{n_2}, \forall j \in [N]. \tag{20c}$$

(ii) If $\Xi = \{0,1\}^{m_1} \times \mathbb{R}^{m_2}$ and the polyhedron $\left\{(\boldsymbol{\pi}, \boldsymbol{\xi}_q) \in \mathbb{R}_+^{\ell} \times [0,1]^{m_1} : \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q}\right\}$ is integral, then RTSP (1) is equivalent to

$$\begin{aligned} v^* = \min_{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\sigma}} \quad & \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{N} \sum_{j \in [N]} [(\boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q})^\top \boldsymbol{y}^j + \mathbb{I}(\theta > 1) \boldsymbol{e}^\top (\boldsymbol{Q}^\top \boldsymbol{y}^j)_+], \\ \text{s.t.} \quad & \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j + \boldsymbol{W}\boldsymbol{y}^j - \theta |\boldsymbol{T}(\boldsymbol{x})| \boldsymbol{e} \geq \boldsymbol{h}(\boldsymbol{x}), \forall j \in [N], \\ & \boldsymbol{\sigma}^j \geq \boldsymbol{Q}^\top \boldsymbol{y}^j, \forall j \in [N], \\ & \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y}^j \in \mathbb{R}^{n_2}, \forall j \in [N]. \end{aligned}$$

We next illustrate the proposed formulation (20) using Example 1, where we realize the fact that support of disruption risks is binary, i.e., $\tilde{\boldsymbol{\delta}} \in \{0,1\}^{n_1}$.

Example 3. Following the notation in Example 1, let us consider DR-RFLP with both demand and disruption uncertainties. We further suppose that the reference distance is $\|\cdot\|_\infty$ and the support of $\tilde{\boldsymbol{\xi}}$ is $\{0,1\}^{n_1} \times \mathbb{R}^\ell$. Since the coefficients of uncertain parameters in the constraints (3b) have the same sign, according to Proposition 7, DR-RFLP can be equivalently formulated as the following MILP:

$$v^* = \min_{\boldsymbol{x},\boldsymbol{y}} \ \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{N} \sum_{j\in[N]} \sum_{t\in[\ell]} \sum_{s\in[n_1+1]} \widehat{c}_{ts}(\widehat{d}_t^j + \theta) y_{ts}^j, \tag{21a}$$

$$\text{s.t.} \quad \sum_{s\in[n_1+1]} y_{ts}^j = 1, \forall j \in [N], \forall t \in [\ell], \tag{21b}$$

$$y_{ts}^j \leq \mathbb{I}(\theta < 1)\widehat{\delta}_s^j x_s, \forall j \in [N], \forall t \in [\ell], \forall s \in [n_1], \tag{21c}$$

$$\boldsymbol{x} \in \{0,1\}^{n_1}, \boldsymbol{y}^j \in \mathbb{R}_+^{\ell\times n_1}, \forall j \in [N]. \tag{21d}$$

Clearly, formulation (21) is less conservative than (11), since the right-hand sides of constraints (21c) are no smaller than those in (11c). This demonstrates that exploring binary support can indeed help reduce the conservatism of the distributionally robust models. □

4.2. Tractable Reformulation II: With Objective Uncertainty Only

Unlike Theorem 2, in general, we cannot provide tractable reformulations for the problem (1) with only binary objective uncertainty, and its complexity analysis is postponed to Section 4.4. Instead, we provide a special case where the tractable reformulation can be derived.

Theorem 5. *Suppose that $\Xi = \{0,1\}^{m_1} \times \{\boldsymbol{\xi}_T\}$ and the polyhedron*

$$\left\{(\boldsymbol{\pi}, \boldsymbol{\xi}_q) \in \mathbb{R}_+^\ell \times [0,1]^{m_1} : \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q}, \sum_{t\in\mathcal{C}_0(\boldsymbol{\zeta}_q^j)} \xi_{qt} + \sum_{t\in\mathcal{C}_1(\boldsymbol{\zeta}_q^j)} (1-\xi_{qt}) \leq \kappa \right\}$$

is integral for all $j \in [N]$ and integer $\kappa \in \mathbb{Z}_+$, where sets $\mathcal{C}_0(\boldsymbol{\zeta}_q^j) := \{t \in [m_1] : \zeta_{qt}^j = 0\}$ and $\mathcal{C}_1(\boldsymbol{\zeta}_q^j) := \{t \in [m_1] : \zeta_{qt}^j = 1\}$. Then for any $p \in [1,\infty)$, the function $\mathcal{Z}(\boldsymbol{x})$ is equivalent to

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j\in[N]} \min_{\boldsymbol{y}\in\mathbb{R}^{n_2},\lambda\in\mathbb{R}_+,\boldsymbol{\sigma}\in\mathbb{R}_+^{m_1}} \left\{ (\boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q})^\top \boldsymbol{y} + \lfloor \theta^p \rfloor \lambda + \boldsymbol{e}^\top \boldsymbol{\sigma} : \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{x}), \right.$$
$$\left. \lambda + \sigma_t \geq (\boldsymbol{Q}^\top \boldsymbol{y})_t, \forall t \in \mathcal{C}_0(\boldsymbol{\zeta}_q^j), \lambda + \sigma_t \geq -(\boldsymbol{Q}^\top \boldsymbol{y})_t, \forall t \in \mathcal{C}_1(\boldsymbol{\zeta}_q^j) \right\}. \tag{22}$$

Proof: Since $p \in [1,\infty)$ and $\Xi = \{0,1\}^{m_1} \times \{\boldsymbol{\xi}_T\}$, thus (6) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j\in[N]} \sup_{\boldsymbol{\pi}\in\mathbb{R}_+^\ell,\boldsymbol{\xi}_q\in\{0,1\}^{m_1}} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T)^\top \boldsymbol{\pi} : \|\boldsymbol{\xi}_q - \boldsymbol{\zeta}_q^j\|_p \leq \theta, \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \right\}. \tag{23a}$$

Since both $\boldsymbol{\xi}_q, \boldsymbol{\zeta}_q^j \in \{0,1\}^{m_1}$, let sets $\mathcal{C}_0(\boldsymbol{\zeta}_q^j) := \{t \in [m_1] : \zeta_{qt}^j = 0\}$ and $\mathcal{C}_1(\boldsymbol{\zeta}_q^j) := \{t \in [m_1] : \zeta_{qt}^j = 1\}$. Therefore, we have the following linearization results:

$$\|\boldsymbol{\xi}_q - \boldsymbol{\zeta}_q^j\|_p^p = \sum_{t\in\mathcal{C}_0(\boldsymbol{\zeta}_q^j)} \xi_{qt} + \sum_{t\in\mathcal{C}_1(\boldsymbol{\zeta}_q^j)} (1-\xi_{qt}). \tag{23b}$$

Thus, (23a) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \sup_{\boldsymbol{\pi}\in\mathbb{R}_+^\ell, \boldsymbol{\xi}_q\in\{0,1\}^{m_1}} \Bigg\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T)^\top \boldsymbol{\pi} : \sum_{t\in\mathcal{C}_0(\boldsymbol{\zeta}_q^j)} \xi_{qt} + \sum_{t\in\mathcal{C}_1(\boldsymbol{\zeta}_q^j)} (1-\xi_{qt}) \leq \lfloor \theta^p \rfloor, \boldsymbol{W}^\top\boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \Bigg\}. \tag{23c}$$

Since the constraint system of the inner supremum (23c) is integral according to our assumption, thus, we can relax the binary variables to be continuous. Thus, we have

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \sup_{\boldsymbol{\pi}\in\mathbb{R}_+^\ell, \boldsymbol{\xi}_q\in[0,1]^{m_1}} \Bigg\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T)^\top \boldsymbol{\pi} : \sum_{t\in\mathcal{C}_0(\boldsymbol{\zeta}_q^j)} \xi_{qt} + \sum_{t\in\mathcal{C}_1(\boldsymbol{\zeta}_q^j)} (1-\xi_{qt}) \leq \lfloor \theta^p \rfloor, \boldsymbol{W}^\top\boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{Q}(\boldsymbol{\xi}_q - \boldsymbol{\zeta}_q^j) + \boldsymbol{q} \Bigg\}. \tag{23d}$$

Let $\boldsymbol{y}$ denote the dual variables of the constraints $\boldsymbol{W}^\top\boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{Q}(\boldsymbol{\xi}_q - \boldsymbol{\zeta}_q^j) + \boldsymbol{q}$, λ be the dual variable of constraint $\sum_{t\in\mathcal{C}_0(\boldsymbol{\zeta}_q^j)} \xi_{qt} + \sum_{t\in\mathcal{C}_1(\boldsymbol{\zeta}_q^j)} (1-\xi_{qt}) \leq \lfloor \theta^p \rfloor$, and $\boldsymbol{\sigma}$ be the dual variables of constraints $\boldsymbol{\xi}_q \leq \boldsymbol{e}$. Then according to the strong duality of linear programming due to sufficiently expensive recourse assumption, (23d) is equivalent to (22). □

We make the following remarks about Theorem 5 and its corresponding formulation (22).

(i) Clearly, since problem (1) with objective uncertainty only is a special case of general problem (1), thus the result of Theorem 4 directly follows and is not listed here;

(ii) The penalty term $\lfloor \theta^p \rfloor\lambda + \boldsymbol{e}^\top\boldsymbol{\sigma}$ with auxiliary variables $\lambda, \boldsymbol{\delta}$ is used to enforce the robustness of the formulation. This penalty term becomes

$$\sum_{t\in\mathcal{C}_0(\boldsymbol{\zeta}_q^j)} ((\boldsymbol{Q}^\top\boldsymbol{y})_t)_+ + \sum_{t\in\mathcal{C}_1(\boldsymbol{\zeta}_q^j)} (-(\boldsymbol{Q}^\top\boldsymbol{y})_t)_+$$

if $\theta^p \geq m_1$; and

(iii) If the integrality assumption of the polyhedra in Theorem 5 does not hold, then (22) provides an upper bound for the function $\mathcal{Z}(\boldsymbol{x})$ and this upper bound will become exact when $\theta \to 0$.

According to the representation results in Theorem 5, we provide the following equivalent deterministic reformulation of problem (1).

Proposition 8. *Suppose that* $\Xi = \{0,1\}^{m_1} \times \{\boldsymbol{\xi}_T\}$ *and the polyhedron*

$$\left\{ (\boldsymbol{\pi}, \boldsymbol{\xi}_q) \in \mathbb{R}_+^\ell \times [0,1]^{m_1} : \boldsymbol{W}^\top\boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q}, \sum_{t\in\mathcal{C}_0(\boldsymbol{\zeta}_q^j)} \xi_{qt} + \sum_{t\in\mathcal{C}_1(\boldsymbol{\zeta}_q^j)} (1-\xi_{qt}) \leq \kappa \right\}$$

is integral for all $j\in[N]$ *and integer* $\kappa\in\mathbb{Z}_+$, *where sets* $\mathcal{C}_0(\boldsymbol{\zeta}_q^j) := \{t\in[m_1] : \zeta_{qt}^j = 0\}$ *and* $\mathcal{C}_1(\boldsymbol{\zeta}_q^j) := \{t\in[m_1] : \zeta_{qt}^j = 1\}$. *Then for any* $p\in[1,\infty)$, *problem* (1) *is equivalent to*

$$\begin{aligned} v^* = \min_{\boldsymbol{x},\boldsymbol{y}} \quad & \boldsymbol{c}^\top\boldsymbol{x} + \frac{1}{N}\sum_{j\in[N]} [(\boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q})^\top\boldsymbol{y}^j + \lfloor \theta^p \rfloor\lambda^j + \boldsymbol{e}^\top\boldsymbol{\sigma}^j], \\ \textit{s.t.} \quad & \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T + \boldsymbol{W}\boldsymbol{y}^j \geq \boldsymbol{h}(\boldsymbol{x}), \forall j\in[N], \\ & \lambda^j + \sigma_t^j \geq (\boldsymbol{Q}^\top\boldsymbol{y}^j)_t, \forall j\in[N], \forall t\in\mathcal{C}_0(\boldsymbol{\zeta}_q^j), \\ & \lambda^j + \sigma_t^j \geq -(\boldsymbol{Q}^\top\boldsymbol{y}^j)_t, \forall j\in[N], \forall t\in\mathcal{C}_1(\boldsymbol{\zeta}_q^j), \end{aligned}$$

$$\boldsymbol{x} \in \mathcal{X}, \boldsymbol{y}^j \in \mathbb{R}^{n_2}, \lambda^j, \boldsymbol{\sigma}^j \in \mathbb{R}^{m_1}, \forall j \in [N].$$

4.3. Tractable Reformulation III: With Constraint Uncertainty Only

Similarly, we provide special cases of problem (1) with only binary constraint uncertainty such that the tractable reformulations can be derived.

Theorem 6. *Suppose that $\Xi = \{\boldsymbol{\xi}_q\} \times \{0,1\}^{m_2}$, $p \in [1, \infty)$, and $\theta \in [1, \sqrt[p]{2})$. Then the function $\mathcal{Z}(\boldsymbol{x})$ is equivalent to*

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j \in [N]} \max_{i \in [m_2+1]} \min_{\boldsymbol{y} \in \mathbb{R}^{n_2}} \left\{ (\boldsymbol{Q}\boldsymbol{\xi}_q^j + \boldsymbol{q})^\top \boldsymbol{y} : \boldsymbol{T}(\boldsymbol{x})\widehat{\boldsymbol{\zeta}}_T^{ij} + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{x}) \right\}, \tag{24}$$

where for each $i \in [m_2+1]$ and $\widehat{\boldsymbol{\zeta}}_T^{ij} = \boldsymbol{\zeta}_T^j + \begin{cases} \boldsymbol{0}, & \text{if } i = m_2+1 \\ \boldsymbol{e}_i, & \text{if } i \in \mathcal{C}_0(\boldsymbol{\zeta}_T^j) \\ -\boldsymbol{e}_i, & \text{if } i \in \mathcal{C}_1(\boldsymbol{\zeta}_T^j) \end{cases}$ *, and sets $\mathcal{C}_0(\boldsymbol{\zeta}_T^j) := \{t \in [m_2] : \zeta_{Tt}^j = 0\}$ and $\mathcal{C}_1(\boldsymbol{\zeta}_T^j) := \{t \in [m_2] : \zeta_{Tt}^j = 1\}$.*

Proof: See Appendix A.4. □

We make the following remarks about Theorem 6 and its corresponding formulation (24).

(i) To evaluate the function $\mathcal{Z}(\boldsymbol{x})$, one needs to solve m_1+1 linear programs for each $j \in [N]$;

(ii) If $\theta \in [0,1)$, then according to the proof of Theorem 6,
$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j \in [N]} \min_{\boldsymbol{y} \in \mathbb{R}^{n_2}} \left\{ (\boldsymbol{Q}\boldsymbol{\xi}_q^j + \boldsymbol{q})^\top \boldsymbol{y} : \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{x}) \right\},$$
i.e., the function $\mathcal{Z}(\boldsymbol{x})$ is equivalent to its sampling average approximation counterpart.

Below provides an equivalent deterministic reformulation of problem (1).

Proposition 9. *Suppose that $\Xi = \{\boldsymbol{\xi}_q\} \times \{0,1\}^{m_2}$, $p \in [1, \infty)$, and $\theta \in [1, \sqrt[p]{2})$. Then problem* (1) *is equivalent to*

$$\begin{aligned} v^* = \min_{\boldsymbol{x}, \boldsymbol{\eta}} \quad & \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{N} \sum_{j \in [N]} \eta_j, \\ \text{s.t.} \quad & \eta_j \geq (\boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q})^\top \boldsymbol{y}^{ij}, \forall j \in [N], \forall i \in [m_2+1], \\ & \boldsymbol{T}(\boldsymbol{x})\widehat{\boldsymbol{\zeta}}_T^{ij} + \boldsymbol{W}\boldsymbol{y}^{ij} \geq \boldsymbol{h}(\boldsymbol{x}), \forall j \in [N], \forall i \in [m_2+1], \\ & \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y}^{ij} \in \mathbb{R}^{n_2}, \forall j \in [N], i \in [m_2+1], \end{aligned}$$

where $\{\widehat{\boldsymbol{\zeta}}_T^{ij}\}_{i \in [m_2+1], j \in [N]}$ are defined in Theorem 6.

We note that if the number of the extreme points of dual constraint system of (2) is small, then equivalently, we can represent the recourse function in the form of piece-wise max of affine functions in the random parameters, and the tractable reformulation can be extended to the case with any reference distance $\|\cdot\|_p$ such that $p \in [1, \infty)$.

Proposition 10. *Suppose that $\Xi = \{0,1\}^\tau$, $p \in [1, \infty)$, and $z(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[\max_{i \in [m]}\{\boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + d_i(\boldsymbol{x})\}]$ with affine functions $\boldsymbol{a}_i(\boldsymbol{x}) : \mathbb{R}^{n_1} \to \mathbb{R}^\tau$ and $d_i(\boldsymbol{x}) : \mathbb{R}^{n_1} \to \mathbb{R}$ for each $i \in [m]$. Then*

- *Function $z(\boldsymbol{x})$ is equivalent to*

$$z(\boldsymbol{x}) = \frac{1}{N} \sum_{j \in [N]} \max_{i \in [m]} \sup_{\boldsymbol{\xi} \in [0,1]^\tau} \left\{ \boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\xi} + d_i(\boldsymbol{x}) : \sum_{t \in \mathcal{C}_0(\boldsymbol{\zeta}^j)} \xi_t + \sum_{t \in \mathcal{C}_1(\boldsymbol{\zeta}^j)} (1 - \xi_t) \le \lfloor \theta^p \rfloor \right\}, \tag{25}$$

where sets $\mathcal{C}_0(\boldsymbol{\zeta}^j) := \{t \in [\tau] : \zeta_t^j = 0\}$ and $\mathcal{C}_1(\boldsymbol{\zeta}^j) := \{t \in [\tau] : \zeta_t^j = 1\}$; and

- *Problem* (1) *is equivalent to*

$$v^* = \min_{\boldsymbol{x}, \boldsymbol{\eta}, \lambda, \boldsymbol{\sigma}} \quad \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{N} \sum_{j \in [N]} \eta_j, \tag{26a}$$

$$\text{s.t.} \quad \eta_j \ge \lambda^{ij} \lfloor \theta^p \rfloor + \boldsymbol{a}_i(\boldsymbol{x})^\top \boldsymbol{\zeta}^j + d_i(\boldsymbol{x}), \forall j \in [N], \forall i \in [m], \tag{26b}$$

$$\lambda^{ij} + \sigma_t^{ij} \ge a_{it}(\boldsymbol{x}), \forall j \in [N], \forall i \in [m], \forall t \in \mathcal{C}_0(\boldsymbol{\zeta}^j), \tag{26c}$$

$$\lambda^{ij} + \sigma_t^{ij} \ge -a_{it}(\boldsymbol{x}), \forall j \in [N], \forall i \in [m], \forall t \in \mathcal{C}_1(\boldsymbol{\zeta}^j), \tag{26d}$$

$$\boldsymbol{x} \in \mathcal{X}, \lambda^{ij} \in \mathbb{R}_+, \boldsymbol{\sigma}^{ij} \in \mathbb{R}_+^\tau, \forall j \in [N], \forall i \in [m]. \tag{26e}$$

Proof: See Appendix A.5. □

We will illustrate the proposed formulation in Proposition 9 using Example 1, where we consider that there is no demand uncertainty, i.e., the only uncertain parameters are facility disruptions, and the support of random disruptions is $\{0,1\}^{n_1}$.

Example 4. Following the notation in Example 1, let us consider DR-RFLP with only disruption risks, i.e., the demand is deterministic satisfying $\mathbb{P}\{\tilde{\boldsymbol{d}} = \boldsymbol{d}\} = 1$.

Suppose the reference distance is $\|\cdot\|_1$, the support of $\tilde{\boldsymbol{\xi}}$ is $\{0,1\}^{n_1} \times \{\boldsymbol{d}\}$, and the Wasserstein radius $\theta \in [1, \sqrt[p]{2})$. According to Proposition 9, DR-FRLP with disruption risks can be equivalently formulated as the following MILP:

$$v^* = \min_{\boldsymbol{x}, \boldsymbol{y}} \; \boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{N} \sum_{j \in [N]} \eta_j, \tag{27a}$$

$$\text{s.t.} \; \eta_j \ge \sum_{t \in [\ell]} \sum_{s \in [n_1+1]} \widehat{c}_{ts} d_t y_{ts}^{ij}, \forall j \in [N], \forall i \in [n_1 + 1], \tag{27b}$$

$$\sum_{s \in [n_1+1]} y_{ts}^{ij} = 1, \forall j \in [N], \forall t \in [\ell], \forall i \in [n_1 + 1], \tag{27c}$$

$$y_{ts}^{ij} \le \bar{\delta}_s^{ij} x_s, \forall j \in [N], \forall t \in [\ell], \forall i \in [n_1 + 1], \forall s \in [n_1], \tag{27d}$$

$$\boldsymbol{x} \in \{0,1\}^{n_1}, \boldsymbol{y}^{ij} \in \mathbb{R}_+^{\ell \times n_1}, \forall j \in [N], \forall i \in [n_1 + 1], . \tag{27e}$$

where for each $i \in [n_1 + 1]$ and $\bar{\boldsymbol{\delta}}_T^{ij} = \widehat{\boldsymbol{\delta}}^j + \begin{cases} \boldsymbol{0}, & \text{if } i = n_1 + 1 \\ \boldsymbol{e}_i, & \text{if } i \in \mathcal{C}_0(\widehat{\boldsymbol{\delta}}^j) \\ -\boldsymbol{e}_i, & \text{if } i \in \mathcal{C}_1(\widehat{\boldsymbol{\delta}}^j) \end{cases}$. □

4.4. Complexity Analysis

Finally, we close this section by showing that for general reference distance $\|\cdot\|_p$ with $p \in [1, \infty]$, either with objective uncertainty only or with constraint uncertainty only, computing the function $\mathcal{Z}(\boldsymbol{x})$ with

$N = 1$ can be NP-hard.

Proposition 11. *Computing $\mathcal{Z}(\boldsymbol{x})$ is NP-hard for any $p \in [1, \infty]$ whenever*

(i) (Without Constraint Uncertainty) $N = 1$, $\Xi = \{0,1\}^{m_1} \times \{\boldsymbol{\xi}_T\}$, $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}, \boldsymbol{T}(\boldsymbol{x}) = \mathbf{0}$, and Wasserstein radius $\theta \geq \sqrt[p]{m_1}$; or

(ii) (Without Objective Uncertainty) $N = 1$, $\Xi = \{\boldsymbol{\xi}_q\} \times \{0,1\}^{m_2}$, $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}, \boldsymbol{T}(\boldsymbol{x}) =$ const., and Wasserstein radius $\theta \geq \sqrt[p]{m_2}$.

Proof: See Appendix A.6. □

The results in Proposition 11 clearly imply that computing the optimal value of problem (1) is also NP-hard.

Corollary 1. *Computing the optimal value of problem* (1) *is NP-hard for any $p \in [1, \infty]$ whenever*

(i) (Without Constraint Uncertainty) $N = 1$, $\Xi = \{0,1\}^{m_1} \times \{\boldsymbol{\xi}_T\}$, $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}, \boldsymbol{T}(\boldsymbol{x}) = \mathbf{0}$, and Wasserstein radius $\theta \geq \sqrt[p]{m_1}$; or

(ii) (Without Objective Uncertainty) $N = 1$, $\Xi = \{\boldsymbol{\xi}_q\} \times \{0,1\}^{m_2}$, $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}, \boldsymbol{T}(\boldsymbol{x}) =$ const., and Wasserstein radius $\theta \geq \sqrt[p]{m_2}$.

5. Numerical Illustration

In this section, we present a numerical study to demonstrate the effectiveness of the proposed formulations and also show how to use cross-validation to choose a proper Wasserstein radius θ.

For the demonstration purpose, we studied two models, i.e, Model (11) and Model (21) from Example 1 and Example 3, respectively. We used normalized 49-node instances provided in [15], and thus in these two models, $\ell = n_1 = 49$. The fixed cost and coordinates of candidate locations can be found at the following link `https://drive.google.com/file/d/11-oc9xX2-tTlSxkNuZhZ-qZlo7xQq80J/view?usp=sharing`. We assumed that disruption happens independently and each location has a probability of $p \in \{0.01, 0.05\}$ to be disrupted, i.e., $\mathbb{P}\{\tilde{\delta}_i = 0\} = p$ and $\mathbb{P}\{\tilde{\delta}_i = 1\} = 1 - p$. To ensure the consistency between random vectors $\tilde{\boldsymbol{\delta}}$ and $\tilde{\boldsymbol{d}}$, we normalized $\tilde{\boldsymbol{d}}$ such that for each $t \in [\ell]$ follows i.i.d uniform distribution in the range between 0.05 and 1.0. We also computed the unit transportation cost

$$\widehat{c}_{ts} = 100 \times \text{Euclidean distance between locations } t \in [\ell] \text{ and } j \in [n_1].$$

Finally, for the emergency facility (i.e., dummy facility), we assumed that its unit transportation cost is $M = 10,000$.

To test these two models, we generate $N = 100$ samples of $(\tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{d}})$, where the computational results are displayed in Table 1. In Table 1, the Wasserstein radius θ varies from 0 to 0.18, where $\theta = 0$, both models are reduced to their sampling average approximation counterpart (SAA model) and for each model, we use Opt.Val., Time, and Built Facilities to denote optimal values, computational time, and built facilities output by the model, respectively. To evaluate the robustness of the solution and choose a proper Wasserstein radius, we generated 100 additional samples, evaluated their corresponding objective function values, and computed the 95% confidence intervals of their mean values, which are displayed in the columns titled "Confidence Interval". All the tested instances were executed on a MacBook Pro with a 2.80 GHz processor and 16GB RAM with a call of the commercial solver Gurobi (version 7.5, with default settings).

From Table 1, we see that all the instances can be solved to the optimality within 1 minute, where Model (21) takes a slightly shorter time. We see that when $\theta = 0$, the SAA model underestimates the cost, where the underestimation mainly comes from the expected transportation cost (i.e., wait-and-see cost). When the Wasserstein radius θ increases, the total cost of both Model (11) and Model (21) increase. However, it is seen that for the same $\theta > 0$, the total cost of Model (21) is significantly smaller than that of Model (11). This demonstrates that exploring support information of random parameters can help reduce the risk of distributional uncertainty. In addition, we also see that the set of built facilities of Model (21) does not change when θ grows to 0.16. This demonstrates that the first-stage results from SAA can be robust. When the probability of disruptions P_d increases from 0.01 to 0.05, we see that Model (11) does not allow us to build any facility due to disruptions when $\theta > 0$, while Model (21) still works and finds appropriate facility locations. This further demonstrates the less conservatism of Model (21).

To choose a proper Wasserstein radius, we suggest to select the smallest θ such that its corresponding total cost is beyond the confidence interval. For example, when $P_d = 0.01$, the best Wasserstein radii of Model (11) and Model (21) are $\theta = 0.02$, while when $P_d = 0.05$, the best Wasserstein radius of Model (21) are $\theta = 0.06$.

Table 1: Numerical results of Model (11) and Model (21) from Example 1 and Example 3, where $N = 100, \ell = n_1 = 49$.

P_d	θ	Model (11)				Model (21)			
		Opt.Val.	Time	Built Facilities	Confidence Interval	Opt.Val.	Time	Built Facilities	Confidence Interval
0.01	0.00	7288.04	7.78	[4, 25, 31, 35, 45]	[7232.33, 7379.25]	7288.04	7.74	[4, 25, 31, 35, 45]	[7232.33, 7379.25]
	0.02	7998.84	10.72	[13, 16, 25, 31]	[7641.80, 7862.95]	7453.31	7.07		
	0.04	8344.45	18.82			7618.58	7.59		
	0.06	8673.85	17.64	[13, 16, 21, 30, 31]	[7748.27, 7963.70]	7783.85	8.37		
	0.08	8995.13	24.35			7949.12	7.26		
	0.10	9295.40	41.71	[13, 16, 19, 21, 30, 31]	[7986.44, 8123.68]	8113.81	13.71		
	0.12	9568.05	30.45			8279.66	12.70		
	0.14	9846.47	30.16			8444.93	13.38		
	0.16	10130.65	30.44			8610.20	15.21		
	0.18	10420.59	30.60			8769.09	14.53	[4, 21, 30, 31, 35, 45]	[7335.91, 7470.26]
0.05	0.00	7498.95	9.75	[4, 25, 31, 35, 45]	[7550.51, 7946.76]	7498.95	9.90	[4, 25, 31, 35, 45]	[7550.51, 7946.76]
	0.02	—-[1]	0.78	[]	—-	7672.22	8.57		
	0.04	—-	0.81	[]	—-	7845.49	8.04		
	0.06	—-	0.78	[]	—-	8018.76	16.38		
	0.08	—-	1.34	[]	—-	8192.03	18.62		
	0.10	—-	1.33	[]	—-	8365.30	19.51		
	0.12	—-	1.29	[]	—-	8538.57	19.77		
	0.14	—-	1.30	[]	—-	8711.84	21.04		
	0.16	—-	0.85	[]	—-	8885.11	15.68		
	0.18	—-	0.76	[]	—-	9052.87	19.61	[4, 21, 30, 31, 35, 45]	[7547.08, 7864.01]

[1] —- means that all the customers will be served by the emergency facility.

Acknowledgments

The author would like to thank Dr. Zhi Chen (City University of Hong Kong) for his insightful comments about $\infty-$Wasserstein ambiguity set. The author thanks the referee and editors for the helpful comments.

References

References

[1] Shabbir Ahmed. Two-stage stochastic integer programming: A brief introduction. *Wiley Encyclopedia of Operations Research and Management Science*, 2010.

[2] Güzin Bayraksan and David K Love. Data-driven stochastic programming using phi-divergences. In *The Operations Research Revolution*, pages 1–19. INFORMS, 2015.

[3] Aharon Ben-Tal, Dick Den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

[4] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*, volume 28. Princeton University Press, 2009.

[5] Ahron Ben-Tal and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, volume 2. SIAM, 2001.

[6] Dimitris Bertsimas, Xuan Vinh Doan, Karthik Natarajan, and Chung-Piaw Teo. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3):580–602, 2010.

[7] Dimitris Bertsimas, Shimrit Shtern, and Bradley Sturt. A data-driven approach for multi-stage linear optimization. Available at Optimization Online, 2018.

[8] Dimitris Bertsimas, Shimrit Shtern, and Bradley Sturt. Two-stage sample robust optimization. *arXiv preprint arXiv:1907.07142*, 2019.

[9] Dimitris Bertsimas, Melvyn Sim, and Meilin Zhang. Adaptive distributionally robust optimization. *Management Science*, 65(2):604–618, 2018.

[10] Jose Blanchet, Yang Kang, and Karthyek Murthy. Robust wasserstein profile inference and applications to machine learning. *arXiv preprint arXiv:1610.05627*, 2016.

[11] Jose Blanchet and Karthyek Murthy. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600, 2019.

[12] Zhi Chen, Daniel Kuhn, and Wolfram Wiesemann. Data-driven chance constrained programs over wasserstein balls. *arXiv preprint arXiv:1809.00210*, 2018.

[13] Zhi Chen, Melvyn Sim, and Peng Xiong. Robust stochastic optimization. Available at Optimization Online, 2019.

[14] Zhi Chen and Weijun Xie. Sharing the value-at-risk under distributional ambiguity. Available at Optimization Online, 2019.

[15] T. Cui, Y. Ouyang, and Zuo-Jun Max Shen. Reliable facility location under the risk of disruptions. *Operations Research*, 58(4):998–1011, 2010.

[16] Ting Dai and Wei Qiao. Trading wind power in a competitive electricity market using stochastic programing and game theory. *IEEE Transactions on Sustainable Energy*, 4(3):805–815, 2013.

[17] Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[18] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.

[19] Rui Gao, Xi Chen, and Anton J Kleywegt. Wasserstein distributional robustness and regularization in statistical learning. *arXiv preprint arXiv:1712.06050*, 2017.

[20] Rui Gao and Anton J Kleywegt. Distributionally robust stochastic optimization with wasserstein distance. *arXiv preprint arXiv:1604.02199*, 2016.

[21] Clark R Givens, Rae Michael Shortt, et al. A class of wasserstein metrics for probability distributions. *The Michigan Mathematical Journal*, 31(2):231–240, 1984.

[22] Joel Goh and Melvyn Sim. Distributionally robust optimization and its tractable approximations. *Operations research*, 58(4-part-1):902–917, 2010.

[23] Mehdi Golari, Neng Fan, and Jianhui Wang. Two-stage stochastic optimal islanding operations under severe multiple contingencies in power grids. *Electric Power Systems Research*, 114:68–77, 2014.

[24] Vincent Guigues, Anatoli Juditsky, and Arkadi Nemirovski. Non-asymptotic confidence bounds for the optimal value of a stochastic program. *Optimization Methods and Software*, 32(5):1033–1058, 2017.

[25] 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.

[26] Grani A Hanasusanto, Vladimir Roitch, Daniel Kuhn, and Wolfram Wiesemann. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151:35–62, 2015.

[27] Grani A Hanasusanto, Vladimir Roitch, Daniel Kuhn, and Wolfram Wiesemann. Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, 65(3):751–767, 2017.

[28] Zhaolin Hu and L Jeff Hong. Kullback-leibler divergence constrained distributionally robust optimization. Available at `http://www.optimization-online.org/DB_FILE/2012/11/3677.pdf`, 2012.

[29] Zhengyang Hu and Guiping Hu. A two-stage stochastic programming model for lot-sizing and scheduling under uncertainty. *International Journal of Production Economics*, 180:198–207, 2016.

[30] Ruiwei Jiang and Yongpei Guan. Data-driven chance constrained stochastic program. *Mathematical Programming*, 158:291–327, 2016.

[31] Ruiwei Jiang and Yongpei Guan. Risk-averse two-stage stochastic program with distributional ambiguity. *Operations Research*, 66(5):1390–1405, 2018.

[32] S Soner Kara and S Onut. A stochastic optimization approach for paper recycling reverse logistics network design under uncertainty. *International Journal of Environmental Science & Technology*, 7(4):717–730, 2010.

[33] Richard M Karp. Reducibility among combinatorial problems, complexity of computer computations (proc. sympos., ibm thomas j. watson res. center, yorktown heights, ny, 1972). *Ò*, 36:85–103, 1972.

[34] Bowen Li, Ruiwei Jiang, and Johanna L. Mathieu. Ambiguous risk constraints with moment and unimodality information. *Mathematical Programming*, Nov 2017.

[35] M Li, P Guo, SQ Fang, and LD Zhang. An inexact fuzzy parameter two-stage stochastic programming model for irrigation water allocation under uncertainty. *Stochastic environmental research and risk assessment*, 27(6):1441–1452, 2013.

[36] Mengshi Lu, Lun Ran, and Zuo-Jun Max Shen. Reliable facility location design under uncertain correlated disruptions. *Manufacturing & Service Operations Management*, 17(4):445–455, 2015.

[37] Fengqiao Luo and Sanjay Mehrotra. Decomposition algorithm for distributionally robust optimization using wasserstein metric. *arXiv preprint arXiv:1704.03920*, 2017.

[38] Areesh Mittal, Can Gokalp, and Grani A Hanasusanto. Robust quadratic programming with mixed-integer uncertainty. *INFORMS Journal on Computing*, 2019.

[39] Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, Jul 2017.

[40] Karthik Natarajan and Chung-Piaw Teo. On reduced semidefinite programs for second order moment bounds with applications. *Mathematical Programming*, 161(1):487–518, Jan 2017.

[41] H. K. Okyay, Fikri Karaesmen, and Süleyman Özekici. Newsvendor models with dependent random supply and demand. *Optimization Letters*, 8(3):983–999, 2014.

[42] Krzysztof Postek, Aharon Ben-Tal, Dick Den Hertog, and Bertrand Melenberg. Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Operations Research*, 66(3):814–833, 2018.

[43] Hamed Rahimian and Sanjay Mehrotra. Distributionally robust optimization: A review. Available at Optimization Online, 2019.

[44] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 1970.

[45] Walter Rudin et al. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1964.

[46] Alexander Schrijver. *Theory of linear and integer programming*. John Wiley & Sons, 1998.

[47] Soroosh Shafieezadeh-Abadeh, Peyman Mohajerin Mohajerin Esfahani, and Daniel Kuhn. Distributionally robust logistic regression. In *Advances in Neural Information Processing Systems*, pages 1576–1584, 2015.

[48] Soroosh Shafieezadeh-Abadeh, Daniel Kuhn, and Peyman Mohajerin Esfahani. Regularization via mass transportation. *Journal of Machine Learning Research*, 20(103):1–68, 2019.

[49] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on stochastic programming: modeling and theory*, volume 9. SIAM, 2009.

[50] Nicolás Garcia Trillos and Dejan Slepčev. On the rate of convergence of empirical measures in ∞-transportation distance. *Canadian Journal of Mathematics*, 67(6):1358–1383, 2015.

[51] Qianfan Wang, Jianhui Wang, and Yongpei Guan. Stochastic unit commitment with uncertain demand response. *IEEE Transactions on power systems*, 28(1):562–563, 2012.

[52] Wolfram Wiesemann, Daniel Kuhn, and Melvyn Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

[53] Hongyu Wu, Mohammad Shahidehpour, Zuyi Li, and Wei Tian. Chance-constrained day-ahead scheduling in stochastic power system operation. *IEEE Transactions on Power Systems*, 29(4):1583–1591, 2014.

[54] Weijun Xie. On distributionally robust chance constrained program with wasserstein distance. Available at Optimization Online, 2018.

[55] Weijun Xie and Shabbir Ahmed. Distributionally robust chance constrained optimal power flow with renewables: A conic reformulation. *IEEE Transactions on Power Systems*, 33(2):1860–1867, 2018.

[56] Weijun Xie and Shabbir Ahmed. On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM Journal on Optimization*, 28(2):1151–1182, 2018.

[57] Weijun Xie and Shabbir Ahmed. Bicriteria approximation of chance constrained covering problems. *Operations Research (In Press)*, 2019.

[58] Weijun Xie, Jie Zhang, and Shabbir Ahmed. Distributionally robust bottleneck combinatorial problems: Uncertainty quantification and robust decision making. *arXiv preprint arXiv:2003.00630*, 2020.

[59] Jie Zhang, Weijun Xie, and Subhash C Sarin. Multi-product newsvendor problem with customer-driven demand substitution: A stochastic integer program perspective. Available at Optimization Online, 2018.

[60] Y. Zhang, R. Jiang, and S. Shen. Ambiguous chance-constrained binary programs under mean-covariance information. *SIAM Journal on Optimization*, 28(4):2922–2944, 2018.

[61] Chaoyue Zhao and Yongpei Guan. Data-driven risk-averse stochastic optimization with wasserstein metric. *Operations Research Letters*, 46(2):262–267, 2018.

Appendix A. Proofs

A.1 Proof of Theorem 3

Theorem 3. *Suppose that $\Xi = \{\boldsymbol{\xi}_q\} \times \mathbb{R}^{m_2}$ and $p = 1$. Then the function $\mathcal{Z}(\boldsymbol{x})$ is equivalent to*

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j \in [N]} \max_{r \in \{-1,1\}} \max_{i \in [m_2]} \min_{\boldsymbol{y} \in \mathbb{R}^{n_2}} \left\{ (\boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q})^\top \boldsymbol{y} : \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j + \boldsymbol{W}\boldsymbol{y} - \theta r \boldsymbol{T}(\boldsymbol{x}) \boldsymbol{e}_i \geq \boldsymbol{h}(\boldsymbol{x}) \right\}. \tag{15}$$

Proof: Since $\Xi = \{\boldsymbol{\xi}_q\} \times \mathbb{R}^{m_2}$ and $p = 1$, (6) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j \in [N]} \sup_{\boldsymbol{\pi} \in \mathbb{R}_+^\ell, \boldsymbol{\xi}_T} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T)^\top \boldsymbol{\pi} : \|\boldsymbol{\xi}_T - \boldsymbol{\zeta}_T^j\|_1 \leq \theta, \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \right\},$$

Above, optimizing $\boldsymbol{\xi}_T$ involving dual norm of $\|\cdot\|_1$, we have

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j \in [N]} \sup_{\boldsymbol{\pi} \in \mathbb{R}_+^\ell} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j)^\top \boldsymbol{\pi} + \theta \|\boldsymbol{T}(\boldsymbol{x})^\top \boldsymbol{\pi}\|_\infty : \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \right\}. \tag{28}$$

Since

$$\|\boldsymbol{T}(\boldsymbol{x})^\top \boldsymbol{\pi}\|_\infty = \max_{i \in [m_2]} \max\{(\boldsymbol{T}(\boldsymbol{x})^\top \boldsymbol{\pi})_i, -(\boldsymbol{T}(\boldsymbol{x})^\top \boldsymbol{\pi})_i\}$$

thus, (28) is further equivalent to

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N} \sum_{j \in [N]} \max_{r \in \{-1,1\}} \max_{i \in [m_2]} \sup_{\boldsymbol{\pi} \in \mathbb{R}_+^\ell} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j)^\top \boldsymbol{\pi} + \theta r \boldsymbol{e}_i^\top \boldsymbol{T}(\boldsymbol{x})^\top \boldsymbol{\pi} : \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \right\},$$

Taking the dual of inner supremum and using strong duality of linear programming, we arrive at (15). □

A.2 Proof of Proposition 6

Proposition 6. *Computing $\mathcal{Z}(\boldsymbol{x})$ is NP-hard whenever the reference distance is $\|\cdot\|_p$ with any $p \in (1, \infty]$, $N = 1$, $\Xi = \{\boldsymbol{\xi}_q\} \times \mathbb{R}^{m_2}$, $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{\zeta}_T^1 = \boldsymbol{0}$, and Wasserstein radius $\theta > 0$.*

Proof: Let us first consider the NP-complete problem - feasibility problem of a general binary program [33] which asks

> (Feasibility problem of a general binary program) Given a rational matrix $\boldsymbol{A} \in \mathbb{R}^{t_1 \times t_2}$ and a rational vector $\boldsymbol{b} \in \mathbb{R}^{t_1}$, is there exists a binary vector $\boldsymbol{r} \in \{0,1\}^{t_2}$ such that $\boldsymbol{A}\boldsymbol{r} = \boldsymbol{b}$?

In the representation (6) of the function $\mathcal{Z}(\boldsymbol{x})$, let $\ell = 2t_2, n_2 = m_2 = t_1 + t_2$, and $\boldsymbol{T}(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{I}_{t_2} \\ -\boldsymbol{I}_{t_2} \end{bmatrix}, \boldsymbol{W}^\top = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{I}_{t_2} & \boldsymbol{I}_{t_2} \end{bmatrix}, \boldsymbol{Q} = \boldsymbol{0}, \boldsymbol{q} = \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{e} \end{pmatrix}, \boldsymbol{\pi} = \begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{s} \end{pmatrix}$. Since $N = 1$, $\Xi = \{\boldsymbol{\xi}_q\} \times \mathbb{R}^{m_2}, \boldsymbol{\zeta}_T^1 = \boldsymbol{0}$, and $\theta > 0$, according to the proof of Theorem 3, $\mathcal{Z}(\boldsymbol{x})$ becomes

$$\mathcal{Z}(\boldsymbol{x}) = \sup_{\boldsymbol{r} \in \mathbb{R}_+^{t_2}, \boldsymbol{s} \in \mathbb{R}_+^{t_2}} \left\{ \theta \left\| (\boldsymbol{r} - \boldsymbol{s}) \right\|_{p*} : \boldsymbol{A}\boldsymbol{r} = \boldsymbol{b}, \boldsymbol{r} + \boldsymbol{s} = \boldsymbol{e} \right\}.$$

Since $p \in (1, \infty]$ and $p* = \frac{p}{p-1} \in [1, \infty)$, thus clearly, $\mathcal{Z}(\boldsymbol{x}) = \theta \sqrt[p*]{t_2}$ if and only if there exists a binary feasible solution $(\boldsymbol{r}, \boldsymbol{s}) \in \{0,1\}^{t_2} \times \{0,1\}^{t_2}$ such that $\boldsymbol{A}\boldsymbol{r} = \boldsymbol{b}, \boldsymbol{r} + \boldsymbol{s} = \boldsymbol{e}$, i.e., the binary program $\{\boldsymbol{r} \in \{0,1\}^{t_2} : \boldsymbol{A}\boldsymbol{r} = \boldsymbol{b}\}$ is feasible.

□

A.3 Proof of Theorem 4

Theorem 4. *Suppose* $p=\infty$ *and* $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_{+}^{\ell \times m_1}$ *or* $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_{-}^{\ell \times m_1}$.

(i) If $\Xi=\mathbb{R}^{m_1} \times\{0,1\}^{m_2}$, *then the function* $\mathcal{Z}(\boldsymbol{x})$ *is equivalent to*

$$\mathcal{Z}(\boldsymbol{x})=\begin{cases}\frac{1}{N} \sum_{j \in[N]} \min_{\boldsymbol{y} \in \mathbb{R}^{n_2}}\left\{(\boldsymbol{Q} \boldsymbol{\zeta}_q^j+\boldsymbol{q})^{\top} \boldsymbol{y}+\theta\|\boldsymbol{Q}^{\top} \boldsymbol{y}\|_1:-(-\boldsymbol{T}(\boldsymbol{x}))_{+} \boldsymbol{e}+\boldsymbol{W} \boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{x})\right\}, & \text{if } \theta \geq 1 \\ \frac{1}{N} \sum_{j \in[N]} \min_{\boldsymbol{y} \in \mathbb{R}^{n_2}}\left\{(\boldsymbol{Q} \boldsymbol{\zeta}_q^j+\boldsymbol{q})^{\top} \boldsymbol{y}+\theta\|\boldsymbol{Q}^{\top} \boldsymbol{y}\|_1: \boldsymbol{T}(\boldsymbol{x}) \boldsymbol{\zeta}_T^j+\boldsymbol{W} \boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{x})\right\}, & \text{if } \theta<1\end{cases}; \quad (18)$$

(ii) If $\Xi=\{0,1\}^{m_1} \times \mathbb{R}^{m_2}$ *and the polyhedron* $\left\{(\boldsymbol{\pi}, \boldsymbol{\xi}_q) \in \mathbb{R}_{+}^{\ell} \times[0,1]^{m_1}: \boldsymbol{W}^{\top} \boldsymbol{\pi}=\boldsymbol{Q} \boldsymbol{\xi}_q+\boldsymbol{q}\right\}$ *is integral, then the function* $\mathcal{Z}(\boldsymbol{x})$ *is equivalent to*

$$\mathcal{Z}(\boldsymbol{x})=\begin{cases}\frac{1}{N} \sum_{j \in[N]} \min_{\boldsymbol{y} \in \mathbb{R}^{n_2}}\left\{(\boldsymbol{Q} \boldsymbol{\zeta}_q^j+\boldsymbol{q})^{\top} \boldsymbol{y}: \boldsymbol{T}(\boldsymbol{x}) \boldsymbol{\zeta}_T^j+\boldsymbol{W} \boldsymbol{y}-\theta|\boldsymbol{T}(\boldsymbol{x})| \boldsymbol{e} \geq \boldsymbol{h}(\boldsymbol{x})\right\}, & \text{if } \theta \geq 1 \\ \frac{1}{N} \sum_{j \in[N]} \min_{\boldsymbol{y} \in \mathbb{R}^{n_2}}\left\{(\boldsymbol{Q} \boldsymbol{\zeta}_q^j+\boldsymbol{q})^{\top} \boldsymbol{y}+\boldsymbol{e}^{\top}(\boldsymbol{Q}^{\top} \boldsymbol{y})_{+}: \boldsymbol{T}(\boldsymbol{x}) \boldsymbol{\zeta}_T^j+\boldsymbol{W} \boldsymbol{y}-\theta|\boldsymbol{T}(\boldsymbol{x})| \boldsymbol{e} \geq \boldsymbol{h}(\boldsymbol{x})\right\}, & \text{if } \theta<1\end{cases}. \quad (19)$$

Proof: We will split the proof into two parts.

(i) Since $p=\infty$ and $\Xi=\mathbb{R}^{m_1} \times\{0,1\}^{m_2}$, thus (6) becomes

$$\mathcal{Z}(\boldsymbol{x})=\frac{1}{N} \sum_{j \in[N]} \sup_{\boldsymbol{\pi} \in \mathbb{R}_{+}^{\ell}} \sup_{\boldsymbol{\xi}_T \in\{0,1\}^{m_2}, \boldsymbol{\xi}_q}\left\{(\boldsymbol{h}(\boldsymbol{x})-\boldsymbol{T}(\boldsymbol{x}) \boldsymbol{\xi}_T)^{\top} \boldsymbol{\pi}: \|\boldsymbol{\xi}_q-\boldsymbol{\zeta}_q^j\|_{\infty} \leq \theta,\|\boldsymbol{\xi}_T-\boldsymbol{\zeta}_T^j\|_{\infty} \leq \theta, \boldsymbol{W}^{\top} \boldsymbol{\pi}=\boldsymbol{Q} \boldsymbol{\xi}_q+\boldsymbol{q}\right\}. \quad (29a)$$

Above, $\boldsymbol{\xi}_T, \boldsymbol{\zeta}_T^j \in\{0,1\}^{m_2}$ and $\|\boldsymbol{\xi}_T-\boldsymbol{\zeta}_T^j\|_{\infty} \leq \theta$ imply that if $\theta \geq 1$, then $\boldsymbol{\xi}_T \in\{0,1\}^{m_2}$; otherwise, $\boldsymbol{\xi}_T=\boldsymbol{\zeta}_T^j$. Hence, using the assumption that $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_{+}^{\ell \times m_1}$ or $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_{-}^{\ell \times m_1}$, (29a) further reduces to

$$\mathcal{Z}(\boldsymbol{x})=\frac{1}{N} \sum_{j \in[N]} \sup_{\boldsymbol{\pi} \in \mathbb{R}_{+}^{\ell}, \boldsymbol{\xi}_q}\left\{(\boldsymbol{h}(\boldsymbol{x})-\mathbb{I}(\theta<1) \boldsymbol{T}(\boldsymbol{x}) \boldsymbol{\zeta}_T^j+\mathbb{I}(\theta \geq 1)(-\boldsymbol{T}(\boldsymbol{x}))_{+} \boldsymbol{e})^{\top} \boldsymbol{\pi}: \|\boldsymbol{\xi}_q-\boldsymbol{\zeta}_q^j\|_{\infty} \leq \theta, \boldsymbol{W}^{\top} \boldsymbol{\pi}=\boldsymbol{Q} \boldsymbol{\xi}_q+\boldsymbol{q}\right\}.$$

Following the similar linearization and dualization steps in Theorem 4, we arrive at (18).

(ii) Since $p=\infty$ and $\Xi=\{0,1\}^{m_1} \times \mathbb{R}^{m_2}$, thus (6) becomes

$$\mathcal{Z}(\boldsymbol{x})=\frac{1}{N} \sum_{j \in[N]} \sup_{\boldsymbol{\pi} \in \mathbb{R}_{+}^{\ell}} \sup_{\boldsymbol{\xi}_q \in\{0,1\}^{m_1}, \boldsymbol{\xi}_T}\left\{(\boldsymbol{h}(\boldsymbol{x})-\boldsymbol{T}(\boldsymbol{x}) \boldsymbol{\xi}_T)^{\top} \boldsymbol{\pi}: \|\boldsymbol{\xi}_q-\boldsymbol{\zeta}_q^j\|_{\infty} \leq \theta,\|\boldsymbol{\xi}_T-\boldsymbol{\zeta}_T^j\|_{\infty} \leq \theta, \boldsymbol{W}^{\top} \boldsymbol{\pi}=\boldsymbol{Q} \boldsymbol{\xi}_q+\boldsymbol{q}\right\}. \quad (29b)$$

Optimizing over $\boldsymbol{\xi}_T$ and using the assumption that $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_{+}^{\ell \times m_1}$ or $\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}_{-}^{\ell \times m_1}$, (29b) is now equivalent to

$$\mathcal{Z}(\boldsymbol{x})=\frac{1}{N} \sum_{j \in[N]} \sup_{\boldsymbol{\pi} \in \mathbb{R}_{+}^{\ell}} \sup_{\boldsymbol{\xi}_q \in\{0,1\}^{m_1}}\left\{(\boldsymbol{h}(\boldsymbol{x})-\boldsymbol{T}(\boldsymbol{x}) \boldsymbol{\zeta}_T^j)^{\top} \boldsymbol{\pi}+\theta \boldsymbol{e}^{\top}|\boldsymbol{T}(\boldsymbol{x})|^{\top} \boldsymbol{\pi}: \|\boldsymbol{\xi}_q-\boldsymbol{\zeta}_q^j\|_{\infty} \leq \theta, \boldsymbol{W}^{\top} \boldsymbol{\pi}=\boldsymbol{Q} \boldsymbol{\xi}_q+\boldsymbol{q}\right\}. \quad (29c)$$

Above, $\boldsymbol{\xi}_q, \boldsymbol{\zeta}_q^j \in\{0,1\}^{m_1}$ and $\|\boldsymbol{\xi}_q-\boldsymbol{\zeta}_q^j\|_{\infty} \leq \theta$ implies that if $\theta \geq 1$, then $\boldsymbol{\xi}_q \in\{0,1\}^{m_1}$; otherwise, $\boldsymbol{\xi}_q=\boldsymbol{\zeta}_q^j$. Thus, there are two sub-cases.

(a) If $\theta < 1$, then (29c) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \sup_{\boldsymbol{\pi}\in\mathbb{R}^\ell_+} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j)^\top \boldsymbol{\pi} + \theta \boldsymbol{e}^\top |\boldsymbol{T}(\boldsymbol{x})|^\top \boldsymbol{\pi} : \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q} \right\}.$$

Let $\boldsymbol{y}$ denote the dual variables of constraints $\boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q}$. Then according to the strong duality of linear programming due to sufficiently expensive recourse assumption, we arrive at the first part of (19);

(b) If $\theta \geq 1$, then (29c) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \sup_{\boldsymbol{\pi}\in\mathbb{R}^\ell_+, \boldsymbol{\xi}_q\in\{0,1\}^{m_1}} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j)^\top \boldsymbol{\pi} + \theta \boldsymbol{e}^\top |\boldsymbol{T}(\boldsymbol{x})|^\top \boldsymbol{\pi} : \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \right\}, \tag{29d}$$

Since the constraint system in (29d) is assumed to be integral, thus (29d) is equivalent to its continuous relaxation

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \sup_{\boldsymbol{\pi}\in\mathbb{R}^\ell_+, \boldsymbol{\xi}_q\in[0,1]^{m_1}} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\zeta}_T^j)^\top \boldsymbol{\pi} + \theta \boldsymbol{e}^\top |\boldsymbol{T}(\boldsymbol{x})|^\top \boldsymbol{\pi} : \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \right\},$$

Let $\boldsymbol{y}$ denote the dual variables of constraints $\boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\zeta}_q^j + \boldsymbol{q}$. Then according to strong duality of linear programming due to sufficiently expensive recourse assumption, we arrive at the second part of (19).

□

A.4 *Proof of Theorem 6*

Theorem 6. *Suppose that $\Xi = \{\boldsymbol{\xi}_q\} \times \{0,1\}^{m_2}$, $p \in [1, \infty)$, and $\theta \in [1, \sqrt[p]{2})$. Then the function $\mathcal{Z}(\boldsymbol{x})$ is equivalent to*

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \max_{i\in[m_2+1]} \min_{\boldsymbol{y}\in\mathbb{R}^{n_2}} \left\{ (\boldsymbol{Q}\boldsymbol{\xi}_q^j + \boldsymbol{q})^\top \boldsymbol{y} : \boldsymbol{T}(\boldsymbol{x})\widehat{\boldsymbol{\zeta}}_T^{ij} + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{x}) \right\}, \tag{24}$$

where for each $i \in [m_2+1]$ and $\widehat{\boldsymbol{\zeta}}_T^{ij} = \boldsymbol{\zeta}_T^j + \begin{cases} \boldsymbol{0}, & \text{if } i = m_2+1 \\ \boldsymbol{e}_i, & \text{if } i \in \mathcal{C}_0(\boldsymbol{\zeta}_T^j) \\ -\boldsymbol{e}_i, & \text{if } i \in \mathcal{C}_1(\boldsymbol{\zeta}_T^j) \end{cases}$ *, and sets $\mathcal{C}_0(\boldsymbol{\zeta}_T^j) := \{t \in [m_2] : \zeta_{Tt}^j = 0\}$ and $\mathcal{C}_1(\boldsymbol{\zeta}_T^j) := \{t \in [m_2] : \zeta_{Tt}^j = 1\}$.*

Proof: Since $p \in [1,\infty)$ and $\Xi = \{\boldsymbol{\xi}_q\} \times \{0,1\}^{m_2}$, (6) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \sup_{\boldsymbol{\pi}\in\mathbb{R}^\ell_+, \boldsymbol{\xi}_T\in\{0,1\}^{m_2}} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\boldsymbol{\xi}_T)^\top \boldsymbol{\pi} : \|\boldsymbol{\xi}_T - \boldsymbol{\zeta}_T^j\|_p \leq \theta, \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \right\}, \tag{30a}$$

According to (23b), and the fact that $\theta \in [1, \sqrt[p]{2})$, we know that

$$\left\{ \boldsymbol{\xi}_T \in \{0,1\}^{m_2} : \|\boldsymbol{\xi}_T - \boldsymbol{\zeta}_T^j\|_p \leq \theta \right\} = \{\boldsymbol{0}\} \cup \{\boldsymbol{\zeta}_T^j + \boldsymbol{e}_i\}_{i\in\mathcal{C}_0(\boldsymbol{\zeta}_T^j)} \cup \{\boldsymbol{\zeta}_T^j - \boldsymbol{e}_i\}_{i\in\mathcal{C}_1(\boldsymbol{\zeta}_T^j)} := \{\widehat{\boldsymbol{\zeta}}_T^{ij}\}_{i\in[m_2+1]}$$

Hence, optimizing $\boldsymbol{\xi}_T$ first, we arrive at

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \max_{i\in[m_2+1]} \sup_{\boldsymbol{\pi}\in\mathbb{R}^\ell_+} \left\{ (\boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{T}(\boldsymbol{x})\widehat{\boldsymbol{\zeta}}_T^{ij})^\top \boldsymbol{\pi} : \boldsymbol{W}^\top \boldsymbol{\pi} = \boldsymbol{Q}\boldsymbol{\xi}_q + \boldsymbol{q} \right\}, \tag{30b}$$

Taking the dual of inner supremum and using strong duality of linear programming due to sufficiently expensive recourse assumption, we arrive at (24). □

A.5 Proof of Proposition 10

Proposition 10. *Suppose that* $\Xi = \{0,1\}^\tau$, $p \in [1,\infty)$, *and* $z(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[\max_{i\in[m]}\{\boldsymbol{a}_i(\boldsymbol{x})^\top\boldsymbol{\xi} + d_i(\boldsymbol{x})\}]$ *with affine functions* $\boldsymbol{a}_i(\boldsymbol{x}) : \mathbb{R}^{n_1} \to \mathbb{R}^\tau$ *and* $d_i(\boldsymbol{x}) : \mathbb{R}^{n_1} \to \mathbb{R}$ *for each* $i \in [m]$. *Then*

- *Function* $z(\boldsymbol{x})$ *is equivalent to*

$$z(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \max_{i\in[m]} \sup_{\boldsymbol{\xi}\in[0,1]^\tau} \left\{ \boldsymbol{a}_i(\boldsymbol{x})^\top\boldsymbol{\xi} + d_i(\boldsymbol{x}) : \sum_{t\in\mathcal{C}_0(\boldsymbol{\zeta}^j)} \xi_t + \sum_{t\in\mathcal{C}_1(\boldsymbol{\zeta}^j)} (1-\xi_t) \leq \lfloor\theta^p\rfloor \right\}, \tag{25}$$

  where sets $\mathcal{C}_0(\boldsymbol{\zeta}^j) := \{t \in [\tau] : \zeta_t^j = 0\}$ *and* $\mathcal{C}_1(\boldsymbol{\zeta}^j) := \{t \in [\tau] : \zeta_t^j = 1\}$; *and*

- *Problem* (1) *is equivalent to*

$$v^* = \min_{\boldsymbol{x},\boldsymbol{\eta},\lambda,\boldsymbol{\sigma}} \quad \boldsymbol{c}^\top\boldsymbol{x} + \frac{1}{N}\sum_{j\in[N]} \eta_j, \tag{26a}$$

$$\text{s.t.} \quad \eta_j \geq \lambda^{ij}\lfloor\theta^p\rfloor + \boldsymbol{a}_i(\boldsymbol{x})^\top\boldsymbol{\zeta}^j + d_i(\boldsymbol{x}), \forall j \in [N], \forall i \in [m], \tag{26b}$$

$$\lambda^{ij} + \sigma_t^{ij} \geq a_{it}(\boldsymbol{x}), \forall j \in [N], \forall i \in [m], \forall t \in \mathcal{C}_0(\boldsymbol{\zeta}^j), \tag{26c}$$

$$\lambda^{ij} + \sigma_t^{ij} \geq -a_{it}(\boldsymbol{x}), \forall j \in [N], \forall i \in [m], \forall t \in \mathcal{C}_1(\boldsymbol{\zeta}^j), \tag{26d}$$

$$\boldsymbol{x} \in \mathcal{X}, \lambda^{ij} \in \mathbb{R}_+, \boldsymbol{\sigma}^{ij} \in \mathbb{R}_+^\tau, \forall j \in [N], \forall i \in [m]. \tag{26e}$$

Proof: Since $\Xi = \{0,1\}^\tau$ and $Z(\boldsymbol{x},\boldsymbol{\xi}) = \max_{i\in[m]}\{\boldsymbol{a}_i(\boldsymbol{x})^\top\boldsymbol{\xi} + d_i(\boldsymbol{x})\}$, (7a) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \max_{i\in[m]} \sup_{\boldsymbol{\xi}\in\{0,1\}^\tau} \left\{ \boldsymbol{a}_i(\boldsymbol{x})^\top\boldsymbol{\xi} + d_i(\boldsymbol{x}) : \|\boldsymbol{\xi} - \boldsymbol{\zeta}^j\|_p \leq \theta \right\}. \tag{31a}$$

According to (23b), (31a) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \max_{i\in[m]} \sup_{\boldsymbol{\xi}\in\{0,1\}^\tau} \left\{ \boldsymbol{a}_i(\boldsymbol{x})^\top\boldsymbol{\xi} + d_i(\boldsymbol{x}) : \sum_{t\in\mathcal{C}_0(\boldsymbol{\zeta}^j)} \xi_t + \sum_{t\in\mathcal{C}_1(\boldsymbol{\zeta}^j)} (1-\xi_t) \leq \lfloor\theta^p\rfloor \right\}. \tag{31b}$$

Since the feasible region defined by cardinality constraint is integral, thus, we can relax the binary variables in the inner supremum of (31b) to be continuous. Thus, we arrive at (25).

To derive the formulation (26), let us first take the dual of inner supremum with dual variables $\lambda, \boldsymbol{\sigma}$ and use the strong duality of linear programming which holds due to sufficiently expensive recourse assumption. Thus, (25) is equivalent to

$$\mathcal{Z}(\boldsymbol{x}) = \frac{1}{N}\sum_{j\in[N]} \max_{i\in[m]} \min_{\lambda\in\mathbb{R}_+,\boldsymbol{\sigma}\in\mathbb{R}_+^{m_2}} \big\{ \lambda\lfloor\theta^p\rfloor + \boldsymbol{a}_i(\boldsymbol{x})^\top\boldsymbol{\zeta}^j + d_i(\boldsymbol{x}) :$$
$$\lambda + \sigma_t \geq a_{it}(\boldsymbol{x}), \forall t \in \mathcal{C}_0(\boldsymbol{\zeta}^j), \lambda + \sigma_t \geq -a_{it}(\boldsymbol{x}), \forall t \in \mathcal{C}_1(\boldsymbol{\zeta}^j) \big\}.$$

Then the conclusion follows from a straightforward linearization. □

A.6 Proof of Proposition 11

Proposition 11. *Computing* $\mathcal{Z}(\boldsymbol{x})$ *is NP-hard for any* $p \in [1,\infty]$ *whenever*

(i) (Without Constraint Uncertainty) $N = 1$, $\Xi = \{0,1\}^{m_1} \times \{\boldsymbol{\xi}_T\}$, $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{0}$, *and Wasserstein radius* $\theta \geq \sqrt[p]{m_1}$; *or*

(ii) *(Without Objective Uncertainty)* $N = 1$, $\Xi = \{\boldsymbol{\xi}_q\} \times \{0,1\}^{m_2}$, $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}, \boldsymbol{T}(\boldsymbol{x}) =$ *const., and Wasserstein radius* $\theta \geq \sqrt[p]{m_2}$.

Proof: Similar to Proposition 6, we will prove the complexity result by reducing the problem to a well-known NP-complete problem - feasibility problem of a general binary program [33].

Next, we split the proof into two cases- when $\Xi = \{0,1\}^{m_1} \times \{\boldsymbol{\xi}_T\}$ and when $\Xi = \{\boldsymbol{\xi}_T\} \times \{0,1\}^{m_2}$.

(i) When $N = 1, \Xi = \{0,1\}^{m_1} \times \{\boldsymbol{\xi}_T\}$, let $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}, \boldsymbol{T}(\boldsymbol{x}) = \mathbf{0}$, $\boldsymbol{W}^\top = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{I}_{t_2} \end{bmatrix}, \boldsymbol{Q} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{I}_{t_2} \end{bmatrix}, \boldsymbol{q} = \begin{pmatrix} \boldsymbol{b} \\ \mathbf{0} \end{pmatrix}, \boldsymbol{\pi} = \boldsymbol{r}$, and $\ell = t_2, m_1 = t_2, n_2 = t_1 + t_2$. As $\theta \geq \sqrt[p]{m_1}$, thus (6) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \sup_{\boldsymbol{r} \in \mathbb{R}_+^{t_2}, \boldsymbol{\xi}_q} \left\{ 0 : \boldsymbol{\xi}_q \in \{0,1\}^{t_2}, \boldsymbol{A}\boldsymbol{r} = \boldsymbol{b}, \boldsymbol{r} = \boldsymbol{\xi}_q \right\}.$$

Clearly, $\mathcal{Z}(\boldsymbol{x}) = 0$ if and only if the binary program $\{\boldsymbol{r} \in \{0,1\}^{t_2} : \boldsymbol{A}\boldsymbol{r} = \boldsymbol{b}\}$ is feasible.

(ii) When $N = 1, \Xi = \{\boldsymbol{\xi}_q\} \times \{0,1\}^{m_2}$, let $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}, \boldsymbol{T}(\boldsymbol{x}) = [\boldsymbol{e}_{t_2+1} - \boldsymbol{e}_1, \ldots, \boldsymbol{e}_{2t_2} - \boldsymbol{e}_{t_2}, \boldsymbol{e}_1 - \boldsymbol{e}_{t_2+1}, \ldots, \boldsymbol{e}_{t_2} - \boldsymbol{e}_{2t_2}]$, $\boldsymbol{W}^\top = \begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \boldsymbol{I}_{t_2} & \boldsymbol{I}_{t_2} \end{bmatrix}, \boldsymbol{Q} = \mathbf{0}, \boldsymbol{q} = \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{e} \end{pmatrix}, \boldsymbol{\pi} = \begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{s} \end{pmatrix}$, and $\ell = 2t_2, m_2 = 2t_2, n_2 = t_1 + t_2$. As $\theta \geq \sqrt[p]{m_2}$, thus (6) becomes

$$\mathcal{Z}(\boldsymbol{x}) = \sup_{\boldsymbol{r} \in \mathbb{R}_+^{t_2}, \boldsymbol{s} \in \mathbb{R}_+^{t_2}, \boldsymbol{\xi}_q} \left\{ \sum_{i \in [t_2]} (\xi_{qi} - \xi_{q(t_2+i)})(r_i - s_i) : \boldsymbol{\xi}_q \in \{0,1\}^{2t_2}, \boldsymbol{A}\boldsymbol{r} = \boldsymbol{b}, \boldsymbol{r} + \boldsymbol{s} = \boldsymbol{e} \right\},$$

which is equivalent to

$$\mathcal{Z}(\boldsymbol{x}) = \sup_{\boldsymbol{r} \in \mathbb{R}_+^{t_2}, \boldsymbol{s} \in \mathbb{R}_+^{t_2}} \left\{ \sum_{i \in [t_2]} |(r_i - s_i)| : \boldsymbol{A}\boldsymbol{r} = \boldsymbol{b}, \boldsymbol{r} + \boldsymbol{s} = \boldsymbol{e} \right\}.$$

Above, $\mathcal{Z}(\boldsymbol{x}) = t_2$ if and only if there exists a binary vector $(\boldsymbol{r}, \boldsymbol{s}) \in \{0,1\}^{m_1} \times \{0,1\}^{m_1}$ such that $\boldsymbol{A}\boldsymbol{r} = \boldsymbol{b}, \boldsymbol{r} + \boldsymbol{s} = \boldsymbol{e}$. Thus, $\mathcal{Z}(\boldsymbol{x}) = t_2$ if and only if the binary program $\{\boldsymbol{r} \in \{0,1\}^{t_2} : \boldsymbol{A}\boldsymbol{r} = \boldsymbol{b}\}$ is feasible.

□