Distributionally Robust Optimization with Integer Recourse: Convex Reformulations and Critical Recourse Decisions

Yiling Zhang[a], Zhe Li[b], Eojin Han[c]

[a]*Industrial and Systems Engineering, University of Minnesota, Minneapolis, 55455, MN, USA*
[b]*Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, 55455, MN, USA*
[c]*Mendoza College of Business, University of Notre Dame, Notre Dame, 46556, IN, USA*

Abstract

The paper studies distributionally robust optimization models with integer recourse. We develop a unified framework that provides finite tight convex relaxations under conic moment-based ambiguity sets and Wasserstein ambiguity sets. They provide tractable primal representations without relying on sampling or semi-infinite optimization. Beyond tractability, the relaxations offer interpretability that captures the criticality of recourse decisions. Finally, it is worth mentioning that our framework applies to distributionally robust chance constraints with integer recourse, which has received limited attention in the literature. The framework is broadly applicable to expectation and conditional value-at-risk measures. In particular, it accommodates distributionally robust chance constraints with integer recourse, which has received limited attention in the literature.

Keywords: Distributionally Robust Optimization, Integer Recourse, Conic Optimization, Interpretability, Two-stage Chance Constraints

1. Introduction

We consider an optimization problem in the following form:

$$v_0 := \sup_{\mathbb{P}\in\mathcal{D}} \rho_{\mathbb{P}}\left[f(\boldsymbol{\xi})\right], \tag{1}$$

where $f(\boldsymbol{\xi}) := \max_{\mathbf{y}\in\mathcal{Y}} \mathbf{q}(\boldsymbol{\xi})^\top \mathbf{y}$ and $\mathcal{Y} \subset \mathbb{R}^m$ is either a discrete set or a polyhedron. The objective coefficient $\mathbf{q}(\boldsymbol{\xi}) : \mathbb{R}^{I_1} \mapsto \mathcal{R}^m$ is assumed linear in the random variable $\boldsymbol{\xi} \in \mathcal{R}^{I_1}$. Specifically, $\mathbf{q}(\boldsymbol{\xi}) = Q\boldsymbol{\xi} + \mathbf{q}_0$ with $Q \in \mathbb{R}^{I_1\times m}$ and $\mathbf{q}_0 \in \mathbb{R}^m$. In problem (1), the decision is a probability distribution $\mathbb{P}$ that determines the risk measure $\rho_{\mathbb{P}}[\cdot]$ applied to the random variable $\boldsymbol{\xi}$. Here, the ambiguity set $\mathcal{D}$ captures partial information on the unknown underlying true probability distribution. For simplicity, we assume that the set of $\boldsymbol{\xi}$ such that $f(\boldsymbol{\xi})$ has multiple optimal solutions has a support with measure zero.

A key motivation for studying (1) is to characterize worst-case performance on measures $\rho_{\mathbb{P}}$ when only partial information about the underlying probability distribution $\mathbb{P}$ is available. This setting also arises in distributionally robust optimization (DRO) [1, 2], for decision-making under uncertainty that generalizes classical robust optimization [3]. In DRO, uncertainty pertains to the probability distribution rather than to parameters, and the resulting problem seeks decisions that perform best under the worst-case distribution, typically by solving a minimax problem:

$$\min_{\mathbf{x}\in\mathcal{X}} \ \mathbf{c}^\top\mathbf{x} + \sup_{\mathbb{P}\in\mathcal{D}} \rho_{\mathbb{P}}\left[f_{\mathbf{x}}(\boldsymbol{\xi})\right], \tag{2}$$

where the outer minimization is over $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d$ and the inner maximization selects the most adversarial distribution. Given a fixed decision $\mathbf{x}$, problem (1) corresponds to the inner maximization problem by specifying the objective coefficient $\mathbf{q}(\boldsymbol{\xi}) = \mathbf{h}(\boldsymbol{\xi}) + T(\boldsymbol{\xi})\mathbf{x}$, where both $\mathbf{h}(\boldsymbol{\xi})$ and $T(\boldsymbol{\xi})$ are linear in $\boldsymbol{\xi}$. Here, we use $f_{\mathbf{x}}(\boldsymbol{\xi})$ to emphasize the dependence of the recourse function on $\mathbf{x}$.

1.1. Significance of the Recourse Structure and Relevant Literature

Many important classes of problems concern the integer recourse structure of distributionally robust bounds (1) in the context of the minimax formulation (2). Below, we describe three representative examples to illustrate the significance.

Sum of piecewise linear convex functions. A broad class of applications considers $f_{\mathbf{x}}(\boldsymbol{\xi})$ expressed as the sum of L piecewise linear convex functions in both $\mathbf{x}$ and $\boldsymbol{\xi}$: $f_{\mathbf{x}}(\boldsymbol{\xi}) = \sum_{\ell=1}^{L} \max_{k=1,\cdots,K_\ell} c^0_{\ell k}(\boldsymbol{\xi}) + \mathbf{c}_{\ell k}(\boldsymbol{\xi})^\top \mathbf{x}$, which is widely used in multi-item newsvendor problems and related inventory problems [see, e.g., 4, 5]. This can be equivalently written as

$$f_{\mathbf{x}}(\boldsymbol{\xi}) = \sum_{\ell=1}^{L} \max_{\mathbf{y}_\ell \in \{0,1\}^{K_\ell}} \left\{ \left(\mathbf{c}^0_\ell(\boldsymbol{\xi}) + C_\ell(\boldsymbol{\xi})\mathbf{x} \right)^\top \mathbf{y}_\ell : \ \mathbf{e}^\top \mathbf{y}_\ell = 1 \right\} = \max_{\mathbf{y} \in \mathcal{Y}} \left(\mathbf{c}^0(\boldsymbol{\xi})) + C(\boldsymbol{\xi})\mathbf{x} \right)^\top \mathbf{y},$$

where $\mathcal{Y} = \left\{ (\mathbf{y}_1, \cdots, \mathbf{y}_L) : \mathbf{y}_\ell \in \{0,1\}^{K_\ell}, \ \mathbf{e}^\top \mathbf{y}_\ell = 1 \ \forall \ell = 1, \cdots, L \right\}$. Here, $\mathbf{c}^0_\ell(\boldsymbol{\xi}) := [c^0_{\ell k}(\boldsymbol{\xi})]_{k \in [K_\ell]}$, $C_\ell(\boldsymbol{\xi})$ is a matrix whose k-th row is $\mathbf{c}_{\ell k}(\boldsymbol{\xi})^\top$, $\mathbf{c}^0(\boldsymbol{\xi})$, $C(\boldsymbol{\xi})$, and $\mathbf{y}$ are concatenated vectors or matrices of $\mathbf{c}^0_\ell(\boldsymbol{\xi})$, $C_\ell(\boldsymbol{\xi})$, and $\mathbf{y}_\ell$, respectively.

Risk-averse network interdiction problems. Network interdiction problems [6, 7] involve a sequential game between two players: an interdictor and a network user, with opposing objectives. These problems have diverse applications in practice, such as disrupting illicit supply networks [8], and ensuring resilience and security in critical infrastructure [9, 10]. The risk-averse and DRO formulations [11, 12] take the form of (2), where binary $\mathbf{x}$ denotes interdiction decisions and binary $\mathbf{y}$ denotes the network user's decisions of whether to traverse each arc in the network. When an interdiction decision $x_a = 1$ is made for an arc a with a random success probability ξ_a, it affects the network user by imposing a traveling cost penalty $d_a \xi_a x_a$ when the user traverses that arc. The zero-sum structure between the interdictor and the user yields a two-stage DRO problem with binary recourse in the form of (2).

Chance constraints with integer recourse. In system design problems under uncertainty, requiring a feasible solution for every possible outcome can lead to overly conservative or costly solutions. Chance-constrained programming mitigates this by ensuring feasibility with high probability and has wide applications, such as humanitarian relief [13], healthcare operations [14], power systems [15], and supply chain [16]. In many of them, a first-stage decision is followed by recourse actions once uncertainty is realized. A chance-constrained two-stage problem with integer recourse [17, 18] is written as

$$\min_{\mathbf{x} \in \mathcal{X}} \left\{ \mathbf{c}^\top \mathbf{x} : \ \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\left(f_{\mathbf{x}}(\boldsymbol{\xi}) \geq \tau \right) \leq \alpha \right\},$$

where the chance constraint involves a worst-case probability measure as in (1) with $\rho_{\mathbb{P}}$ as a probability measure given a predetermined target τ.

Methodological relevance and contributions. On the methodological side, early work studies analytical bounds for structured objectives under mean, variance, or mean absolute deviation assumptions. More recently, bi-dual reformulation approaches have enabled tractable convex representations in a broader class of problems, including moment-based DRO problems, problems with Wasserstein-distance ambiguity sets, and optimal transport generalizations [2]. Our work contributes from a primal perspective, developing finite convex reformulations that are valid across a range of risk measures, including expectation, probability, and conditional value-at-risk (CVaR), without relying on sampling or semi-infinite programming. This is also in line with a recent work [19] on a primal perspective to identify a discrete subset of the support that determines the worst-case distribution, whereas our work focuses on the interplay between the worst-case distribution and critical recourse decisions.

Beyond tractability, our framework yields interpretable quantities that enable us to identify a subset of recourse actions and quantify their impacts under the worst-case distribution. Furthermore, we extend these results to distributionally robust chance constraints with integer recourse, which, to the best of our knowledge, has not been systematically addressed in the existing literature.

In the rest of the paper, we introduce a second-order cone-based ambiguity set in Section 1.2. Sections 2 and 3 develop finite convex relaxations and their tightness using expectation and probability measures as $\rho_{\mathbb{P}}$ in Problem (1), respectively. We close with numerical results in Section 4. The Electronic Companion contains proofs, a column-constraint algorithm, and extensions to Wasserstein ambiguity sets.

Notation. We use the shorthand $[N] = \{1, \ldots, N\}$ to represent the set of all integers up to N. Collections of vectors $\{\mathbf{y}\}_{k=1}^{K}$ is denoted by $\mathbf{y}^{[K]}$. We use $(\cdot)^+$ to denote $\max\{\cdot, 0\}$. We use $\mathcal{M}(\mathbb{R}^I)$ to represent the set of probability distributions on $\mathbb{R}^I$. We denote $\rho_{\mathbb{P}}[\cdot]$ as the risk measure over the probability distribution $\mathbb{P}$, $\mathbb{E}_{\mathbb{P}}[\cdot]$ as the expectation, and $\mathbb{P}(\cdot)$ as the probability of event $\cdot$.

1.2. The SOC Ambiguity Set

In this paper, we focus on an ambiguity set based on second-order cones [see, e.g., 20, 21]. Our results can be extended to ambiguity sets based on generalized cones and Wasserstein metrics; the latter is presented in Appendix E.

Let g denote a function for $\mathcal{U} \mapsto \mathbb{R}_+^{I_2}$. We consider the partial cross-moment ambiguity set $\mathcal{D}$:

$$\mathcal{D} := \left\{\mathbb{P} \in \mathcal{M}(\mathbb{R}^{I_1}) : \ \mathbb{P}_{\mathbb{P}}(\boldsymbol{\xi} \in \mathcal{U}) = 1, \ \mathbb{E}_{\mathbb{P}}[G\boldsymbol{\xi}] = \boldsymbol{\mu}, \ \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{\xi})] \leq \boldsymbol{\sigma}\right\},$$

where $G \in \mathbb{R}^{L_1 \times I_1}$, $\boldsymbol{\mu} \in \mathbb{R}^{L_1}$, and $\boldsymbol{\sigma} \in \mathbb{R}^{I_2}$. Here, the support set $\mathcal{U} \subset \mathbb{R}^{I_1}$ is second-order cone (SOC) representable, and the epigraph of each $g_i, \ i \in [I_2]$, i.e., $\operatorname{epi}(g_i) := \left\{(\boldsymbol{\xi}, z) \in \mathbb{R}^{I_1+1} : \ g_i(\boldsymbol{\xi}) \leq z\right\}$ is also SOC representable.

The SOC ambiguity set $\mathcal{D}$ can describe a wide variety of useful distributional characterizations, including bounds on absolute deviation, variance, semivariance, and entropy. We refer interested readers to Bertsimas et al. [21] for more distributional characterizations.

Following [20, 21], we introduce a lifting variable $\boldsymbol{\zeta} \in \mathbb{R}^{I_2}$ and consider the following *lifted* ambiguity set

$$\mathcal{G} := \left\{\mathbb{P} \in \mathcal{M}(\mathbb{R}^{I_1+I_2}) : \ \mathbb{P}((\boldsymbol{\xi}, \boldsymbol{\zeta}) \in \bar{\mathcal{U}}) = 1, \ \mathbb{E}_{\mathbb{P}}[G\boldsymbol{\xi}] = \boldsymbol{\mu}, \ \mathbb{E}_{\mathbb{P}}[\boldsymbol{\zeta}] \leq \boldsymbol{\sigma}\right\},$$

where $\bar{\mathcal{U}}$ is the lifted support set defined as

$$\bar{\mathcal{U}} := \left\{(\boldsymbol{\xi}, \boldsymbol{\zeta}) \in \mathbb{R}^{I_1+I_2} : \ \boldsymbol{\xi} \in \mathcal{U}, \ g(\boldsymbol{\xi}) \leq \boldsymbol{\zeta}\right\},$$

with $g(\boldsymbol{\xi}) = (g_1(\boldsymbol{\xi}), \ldots, g_{I_2}(\boldsymbol{\xi}))$. Given the SOC representable support set $\mathcal{U}$ and epigraphs $\operatorname{epi}(g_i), \ i \in [I_2]$, the lifted support set $\bar{\mathcal{U}}$ is also SOC representable. Without loss of generality, we assume that there exist $O \in \mathbb{R}^{L_2 \times I_1}$, $P \in \mathbb{R}^{L_2 \times I_2}$, $S \in \mathbb{R}^{L_2 \times I_3}$, $\mathbf{s} \in \mathbb{R}^{L_2}$, and a Cartesian product of second-order cones $\mathcal{K} \subset \mathbb{R}^{L_2}$ such that $\bar{\mathcal{U}} = \{(\boldsymbol{\xi}, \boldsymbol{\zeta}) \in \mathbb{R}^{I_1+I_2} : \ \exists \mathbf{v} \in \mathbb{R}^{I_3}, \ O\boldsymbol{\xi} + P\boldsymbol{\zeta} + S\mathbf{v} \preccurlyeq_{\mathcal{K}} \mathbf{s}\}$. An important observation is that the ambiguity set $\mathcal{D}$ is equivalent to the set of marginal distributions of $\boldsymbol{\xi}$ under all distributions in the lifted set $\mathcal{G}$ [see Proposition 1, 21], and is formalized in the following corollary.

Corollary 1. *The worst-case bound in Problem* (1) *is equivalent under the lifted ambiguity set, i.e.,* $\sup_{\mathbb{P} \in \mathcal{D}} \rho_{\mathbb{P}}[f(\boldsymbol{\xi})] = \sup_{\mathbb{P} \in \mathcal{G}} \rho_{\mathbb{P}}[f(\boldsymbol{\xi})]$.

In this paper, we make the bounded assumption on the support:

Assumption 1 (Bounded Support)**.** *The support set $\mathcal{U}$ is bounded and there exist element-wise finite upper and lower bounds* $\bar{\boldsymbol{\xi}}_i := \sup_{\boldsymbol{\xi} \in \mathcal{U}} \{\boldsymbol{\xi}_i\}$ *and* $\underline{\boldsymbol{\xi}}_i := \inf_{\boldsymbol{\xi} \in \mathcal{U}} \{\boldsymbol{\xi}_i\}, \ i \in [I_1]$.

This is a mild assumption in data-driven settings where the knowledge of uncertain parameters is restricted to a set of samples drawn independently from the underlying distribution. For instance, Delage and Ye [22] discuss how to construct a bounded confidence region to support the unknown underlying distribution with high probability.

2. Expectation

In this section, we use expectation as the risk measure $\rho_{\mathbb{P}}$ in (1). When all the feasible solutions of the recourse feasible region are known, i.e., $\mathcal{Y} = \{\mathbf{y}^k\}_{k=1}^{K}$, where K is the number of all feasible solutions, which is not necessarily finite. For each realization of $\boldsymbol{\xi} \in \mathcal{U}$, let $\hat{\mathbf{y}}(\boldsymbol{\xi})$ denote the optimal solution of the recourse function $f(\boldsymbol{\xi})$. Since $\boldsymbol{\xi}$ is random, $\hat{\mathbf{y}}(\boldsymbol{\xi})$ is a random variable depending on $\boldsymbol{\xi}$. Given a distribution $\mathbb{P} \in \mathcal{G}$, for $k = 1, \ldots, K$, let

$$u_k = \mathbb{P}(\hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k), \ \mathbf{w}_k = \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi} \mid \hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k]\mathbb{P}(\hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k), \text{ and } \mathbf{z}_k = \mathbb{E}_{\mathbb{P}}[\boldsymbol{\zeta} \mid \hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k]\mathbb{P}(\hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k).$$

Observe that $\sum_{k=1}^{K} u_k = 1$, $G \sum_{k=1}^{K} \mathbf{w}_k = \boldsymbol{\mu}$, and $\sum_{k=1}^{K} \mathbf{z}_k \leq \boldsymbol{\sigma}$. Following the convexity of the lifted support set $\bar{\mathcal{U}}$, we have $O\mathbf{w}_k + P\mathbf{z}_k + S\mathbf{v}_k \preccurlyeq_{\mathcal{K}} \mathbf{s}u_k$ and $\bar{\boldsymbol{\xi}} u_k \geq \mathbf{w}_k \geq \underline{\boldsymbol{\xi}} u_k$ for some $\mathbf{v}_k, \ k \in [K]$.

Due to the convexity of the lifted support set, for positive u_k, we have $(\mathbf{w}_k/u_k, \mathbf{z}_k/u_k) \in \bar{\mathcal{U}}$, implying $O\mathbf{w}_k + P\mathbf{z}_k + S\mathbf{v}_k \preccurlyeq \mathbf{s}u_k$. Following the law of total probability, the expectation $\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{\xi})]$ can be expressed as

$$\mathbb{E}_{\mathbb{P}}[f(\boldsymbol{\xi})] = \sum_{k=1}^{K} \mathbb{E}_{\mathbb{P}}\left[\mathbf{q}(\boldsymbol{\xi})^\top \mathbf{y}^k \,\middle|\, \hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k\right] \mathbb{P}\left(\hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k\right).$$

These observations lead to the following convex second-order cone programming (SOCP) relaxation of the worst-case bound in (1), which is shown to be tight in Theorem 1, with the proof provided in Appendix A.

Theorem 1. *Given all the feasible solutions of $\mathcal{Y} = \{\mathbf{y}^k\}_{k=1}^K$, the worst-case expectation of $f(\xi)$ in* (1) *equals to*

$$\begin{aligned} V(\mathbf{y}^{[K]}) := \max_{(u_k, \mathbf{w}_k, \mathbf{z}_k, \mathbf{v}_k)_{k=1}^K} \quad & \sum_{k=1}^{K} (Q\mathbf{w}_k + \mathbf{q}_0 u_k)^\top \mathbf{y}^k & \text{(3a)} \\ \text{s.t.} \quad & \sum_{k=1}^{K} u_k = 1,\ G\sum_{k=1}^{K} \mathbf{w}_k = \boldsymbol{\mu},\ \sum_{k=1}^{K} \mathbf{z}_k \leq \boldsymbol{\sigma} & \text{(3b)} \\ & O\mathbf{w}_k + P\mathbf{z}_k + S\mathbf{v}_k \preccurlyeq_{\mathcal{K}} \mathbf{s}u_k,\ k \in [K] & \text{(3c)} \\ & \bar{\boldsymbol{\xi}} u_k \geq \mathbf{w}_k \geq \underline{\boldsymbol{\xi}} u_k,\ k \in [K] & \text{(3d)} \\ & u_k \geq \mathbf{0},\ \mathbf{z}_k \geq \mathbf{0},\ k \in [K]. & \text{(3e)} \end{aligned}$$

It is clear that the above formulation constitutes a valid convex relaxation of problem (1). To show the upper bound is tight, we construct a worst-case distribution that achieves the bound, which is presented in Appendix A. If $|\mathcal{Y}|$ is small, the worst-case bound v_0 can be efficiently evaluated by solving the convex SOCP reformulation (3). Even if $|\mathcal{Y}|$ is large or infinite, we can show that only a finite number of $\mathbf{y}$-solutions are needed to construct an exact SOCP reformulation. Specifically, we show in the following theorem that there exists a finite number K_0 of $\mathbf{y}$-solutions, denoted as $\{\mathbf{y}^k\}_{k=1}^{K_0}$, that achieve the same tight bound $V(\mathbf{y}^{[K_0]}) = v_0$. The proof is presented in Appendix B.

Theorem 2 (Finite Convex Reformulation)**.** *There exists a subset $\mathcal{Y}' \subset \mathcal{Y}$ of size $K_0 \leq \min\{I_1 + I_2 + 2, K\}$ such that $V(\mathbf{y}^{[K_0]}) = v_0$.*

Remark (On the number of recourse solutions). In practice, the number of feasible recourse $\mathbf{y}$ can be much smaller depending on the distribution of $\mathbf{q}(\xi)$. Intuitively, if the distribution of $\mathbf{q}(\xi)$ is more concentrated around its nominal value, a subset of very few recourse actions may become active with strictly positive probabilities.

Remark (On enumerating recourse solutions). The results above assume that all feasible recourse solutions $\mathcal{Y} = \{\mathbf{y}^k\}_{k=1}^K$ are known. This assumption is not restrictive when the feasible region of moderate size, for instance, when the recourse function $f(\boldsymbol{\xi})$ is piecewise linear allowing all feasible $\mathbf{y}$ to be identified efficiently in advance. When $\mathcal{Y}$ is too large to enumerate directly, one can start with a small subset of recourse solutions and iteratively add back others as needed. To achieve this, we develop a column-constraint generation method with the details provided in Appendix C.

Remark (On Interpretability and Persistence). An important byproduct of the finite convex reformulation is that their solutions quantify the contribution of each recourse solution $\mathbf{y}^k$ to the overall objective in problem (1). In particular, the variable u_k quantifies the relative criticality of the recourse solution $\mathbf{y}^k$ under the worst-case distribution. This is reminiscent of the persistence analysis in stochastic and robust integer programming [23, 24], where persistence captures the likelihood that a discrete decision remains fixed across realizations of uncertainty. Our results extend this idea by incorporating partial crossmoment and structured support information, leading to less conservative persistence results. In addition, our results provide further statistics, such as expected objective contribution via $\mathbf{w}_k$ and its variance via $\mathbf{v}_k$.

Remark (On Extensions). Although the discussion focuses on the SOC ambiguity set, the results extend naturally to other conic representations. For example, instead of SOC sets, one can use general conic representable sets where $\mathcal{K}$ is relaxed to a Cartesian product of general cones rather than second-order cones. Moreover, similar finite convex reformulations can also be derived for Wasserstein ambiguity sets based on the conditional distribution representation. Further details are available in Appendix E.

3. Probability Bound

In this section, we consider a probability measure in (1): $\rho_{\mathbb{P}}(f(\boldsymbol{\xi})) = \mathbb{P}(f(\boldsymbol{\xi}) \geq \tau)$, where τ is a given target of the recourse objective. Again, we assume that all the feasible solutions of the recourse problem are known, i.e., $\mathcal{Y} = \{\mathbf{y}^k\}_{k=1}^K$, where K is the number of all feasible solutions. For every realization of $\boldsymbol{\xi} \in \mathcal{U}$, let $\hat{\mathbf{y}}(\boldsymbol{\xi})$ denote the optimal solution of the recourse function $f(\boldsymbol{\xi})$ such that $f(\boldsymbol{\xi}) \geq \tau$. Since $\boldsymbol{\xi}$ is random, $\hat{\mathbf{y}}(\boldsymbol{\xi})$ is a random variable depending on $\boldsymbol{\xi}$. Given a distribution $\mathbb{P} \in \mathcal{D}$, for $k = 1, \ldots, K$, let

$$u_k = \mathbb{P}(\hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k),\ \mathbf{w}_k = \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi} \mid \hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k]\mathbb{P}(\hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k),\ \text{and}\ \mathbf{z}_k = \mathbb{E}_{\mathbb{P}}[\boldsymbol{\zeta} \mid \hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k]\mathbb{P}(\hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k).$$

The expectation operator preserves the linear inequality $f(\boldsymbol{\xi}) \geq \tau$, i.e.,

$$\mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{\xi}) \mid \hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k\right] \geq \tau \Leftrightarrow \mathbb{E}_{\mathbb{P}}\left[\mathbf{q}(\boldsymbol{\xi})^\top \hat{\mathbf{y}}^k \,\middle|\, \hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k\right]\mathbb{P}\left(\hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k\right) \geq \tau\mathbb{P}\left(\hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k\right).$$

Moreover, we introduce $u_0 = \mathbb{P}(f(\boldsymbol{\xi}) < \tau)$, $\mathbf{w}_0 = \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi} | f(\boldsymbol{\xi}) < \tau]\mathbb{P}(f(\boldsymbol{\xi}) < \tau)$, and $\mathbf{z}_0 = \mathbb{E}_{\mathbb{P}}[\boldsymbol{\zeta} | f(\boldsymbol{\xi}) < \tau]\mathbb{P}(f(\boldsymbol{\xi}) < \tau)$. Note that $\sum_{k=0}^K u_k = 1$, $G\sum_{k=0}^K \mathbf{w}_k = \boldsymbol{\mu}$, and $\sum_{k=0}^K \mathbf{z}_k \leq \boldsymbol{\sigma}$. Similar to the expectation case, we can derive a tight SOCP relaxation of the worst-case probability bound.

Theorem 3. *Given all the feasible solutions of $\mathcal{Y} = \{\mathbf{y}^k\}_{k=1}^K$, the worst-case probability $\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{P}(f(\boldsymbol{\xi}) \geq \tau)$ is equivalent to*

$$
\begin{aligned}
U(\mathbf{y}^{[K]}) := \max_{(u_k, \mathbf{w}_k, \mathbf{z}_k, \mathbf{v}_k)_{k=0}^K} \quad & \sum_{k=1}^K u_k && \text{(4a)}\\
s.t. \quad & (Q\mathbf{w}_k + q_0 u_k)^\top \mathbf{y}^k \geq \tau u_k,\ k \in [K] && \text{(4b)}\\
& \sum_{k=0}^K u_k = 1,\ G\sum_{k=0}^K \mathbf{w}_k = \boldsymbol{\mu},\ \sum_{k=0}^K \mathbf{z}_k \leq \boldsymbol{\sigma} && \text{(4c)}\\
& O\mathbf{w}_k + P\mathbf{z}_k + Q\mathbf{v}_k \preccurlyeq_{\mathcal{K}} \mathbf{s}u_i,\ k \in [K] \cup \{0\} && \text{(4d)}\\
& \bar{\boldsymbol{\xi}} u_k \geq \mathbf{w}_k \geq \underline{\boldsymbol{\xi}} u_k,\ k \in [K] \cup \{0\} && \text{(4e)}\\
& \mathbf{z}_k \geq 0,\ u_k \geq 0,\ k \in [K] \cup \{0\} && \text{(4f)}
\end{aligned}
$$

The proof of Theorem 3 is presented in Appendix D. Following a similar argument for Theorem 2, only a finite number of $\mathbf{y}$-solutions are needed even if the set $\mathcal{Y}$ contains an infinite number of solutions, which is formally presented in the following.

Corollary 2 (Finite Convex Reformulation)**.** *There exists a subset $\mathcal{Y}' \subset \mathcal{Y}$ of size $K_0 \leq \min\{I_1 + I_2 + 2, K + 1\}$ such that $U(\mathbf{y}^{[K_0]}) = v_0$.*

4. Numerical Studies

In this section, we demonstrate the proposed framework through two examples: a project management problem and a multi-item newsvendor problem. All the computations are implemented in JuMP v1.26.0 (Julia) and solved with Gurobi v1.7.4. Experiments are run on an Apple MacBook Air with an M4 8-core processor and 24 GB of RAM.

Project Management: Value of Incorporating Partial Crossmoments. Following Bertsimas et al. [23] and Van Slyke [25], we study a project management problem with eight activities distributed over five paths, as illustrated in Figure 1. Each arc represents an activity whose expected duration and variance are shown in the figure.

Assuming only marginal distributional information, Bertsimas et al. [23] identify critical paths by computing the probability of each activity path. We extend this analysis to evaluate the value of incorporating partial crossmoment information. Our proposed model assumes known mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}$ for all activities, with an upper bound

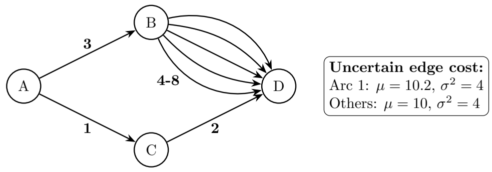

Figure 1: Project network structure.

Methods	Expected Duration	Probability (Criticality)					
		$3\to4$	$3\to5$	$3\to6$	$3\to7$	$3\to8$	$1\to2$
MCS	22.971	0.116	0.118	0.172	0.170	0.138	0.286
MM	26.295	0.131	0.131	0.131	0.131	0.131	0.343
PCM	26.024	0.114	0.114	0.151	0.151	0.134	0.338

Figure 2: Comparison of expected duration and path criticalities.

$\underline{\boldsymbol{\xi}} = \boldsymbol{\mu} + 3\boldsymbol{\sigma}$ and a lower bound $\overline{\boldsymbol{\xi}} = \boldsymbol{\mu} - 3\boldsymbol{\sigma}$. We compare two models: MM using only marginal moments, and PCM additionally incorporating partial crossmoments of the form

$$\mathbb{E}_{\mathbb{P}}[\left(\mathbf{f}_i^\top(\boldsymbol{\xi}-\boldsymbol{\mu})\right)^2] \le \mathbf{f}_i^\top \Sigma \mathbf{f}_i,\ i = 1,\ldots,8,\ \text{where } \mathbf{f}_{ij} = \begin{cases} 2 & \text{if } j = i, \\ -\frac{1}{2} & \text{otherwise} \end{cases},\ i = 1,\ldots,8.$$

Correlations among all activities are set to zero except for activities $(4,5)$ and $(6,7)$, with correlation 0.5 and -0.75, respectively. Without assuming an exact distribution, our model obtains the worst case expected completion time. Additionally, a Monte Carlo Simulation (MCS) is performed using 50,000 samples drawn from a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance Σ.

Table 2 shows the expected duration and probability of each path. The PCM model yields a smaller expected completion time than MM due to the incorporation of partial crossmoment information. While MM assigns equal probabilities to the five paths passing through arc 3, PCM results in different probabilities in better agreement with the MCS results. Our results extend the persistence analysis [23] by incorporating partial crossmoment information.

Multi-item Newsvendor Problem: Comparison Across Risk Levels. We next consider a newsvendor problem with N items. Each item has a unit wholesale price $c_n = \$10$ and retail price $v_n = \$5$, $n \in [N]$. The salvage profit $g_n(x) = \min_{i=1,2,3} \alpha_n^g x_n + \beta_n^g$ is a piecewise concave function with three pieces, while the goodwill cost $b_n(x) = \max_{i=1,2,3} \alpha_n^b x_n + \beta_n^b$ is a piecewise convex function with three pieces. The parameters are specified as $\alpha_n^b = [0.5, 1.0, 1.5]$, $\beta_n^b = [0.0, -1.0, -3.0]$, $\alpha_n^g = [2.5, 2.0, 1.5]$, $\beta_n^g = [0.0, 1.0, 3.0]$. Given an order quantity $\mathbf{x}$, the cost function

$$f_{\mathbf{x}}(\boldsymbol{\xi}) = \sum_{n=1}^{N} \max\left\{-v_n\xi_n - g_n(x_n - \xi_n), -v_n x_n + b_n(\xi_n - x_n)\right\}$$

is a sum of piecewise convex functions, each with six pieces. The recourse function can thus be rewritten in the form of (2) by introducing recourse $\mathbf{y} \in \{0,1\}^{6N}$, where positions $6n-5$ to $6n$ correspond to the six pieces of item $n \in [N]$, with the first three for the salvage function and the remaining for the goodwill cost function.

N	2	3	4	5	6	7
K_0	4	5	7	8	11	13
K	6^2	6^3	6^4	6^5	6^6	6^7
$I_1 + I_2 + 2$	7	9	11	13	15	17

Table 1: Number of $\mathbf{y}$ with positive probabilities.

$\hat{\mathbf{y}}$	$\mathbb{P}(\hat{\mathbf{y}})$	$\mathbb{E}_{\mathbb{P}}[f_{\mathbf{x}^\star}(\boldsymbol{\xi}) \mid \hat{\mathbf{y}}]$	$\mathrm{Var}_{\mathbb{P}}[f_{\mathbf{x}^\star}(\boldsymbol{\xi}) \mid \hat{\mathbf{y}}]$
$\mathbf{y}_{g^1,g^2}$	0.378	-343.87	214.24
$\mathbf{y}_{b^1,g^2}$	0.289	-353.70	173.44
$\mathbf{y}_{g^1,b^2}$	0.309	-360.98	55.15
$\mathbf{y}_{b^1,b^2}$	0.023	-370.92	11.71

Table 2: Solution results with expectation ($\epsilon = 0.0$).

Theorem 2 suggests that the number of distinct $\mathbf{y}$-solutions with positive probability is no more than the number of constraints in the ambiguity set plus one and the cardinality of the feasible region $\mathcal{Y}$. That is, $K_0 \le \min\{K, I_1 + I_2 + 2\}$. Table 1 shows their values for newsvendor problems using expectation with $N = 2,\ldots,7$ items. The actual number of active $\mathbf{y}$-solutions is much smaller than the elements in the feasible region $\mathcal{Y}$.

The convex reformulation results can be directly extended to CVaR of risk level ϵ, which measures the expected loss exceeding the value-at-risk (VaR), the ϵ-percentile [26]. The expectation measure can be viewed as a special case of CVaR when $\epsilon = 0.0$. Next, we examine how critical resource decisions evolve with increasing risk levels in a two-item newsvendor problem. For notation simplicity, we use two subscripts to the status of the two items. For example, $\mathbf{y}_{g^1,g^1}$ represents a $\mathbf{y}$-solution with all zeros, except that $\mathbf{y}_4 = \mathbf{y}_{10} = 1$. That is, both items have overage and the excessive amount triggers their first piece of the salvage functions.

Table 2 reports the probabilities and conditional moments of feasible recourse decisions at risk level ϵ. The dominant solution $\mathbf{y}_{g^1,g^2}$ (overage for both items) yields an expected cost of -343.87. The second dominant solution $\mathbf{y}_{g^1,b^2}$ (overage for item 1 and shortage for item 2) has a smaller expected cost but lower conditional variance, implying more concentrated cost realizations. Figures 3-4 illustrate the partition of the support set and the associated optimal recourse solutions. In Figure 4, the dark colored region represents the realizations of the recourse function below the VaR= -338.08 at $\epsilon = 0.7$. As risk aversion increases, the probability mass assigned to high-cost recourse decisions decreases, indicating to prioritize solutions that mitigate extreme unfavorable outcomes.

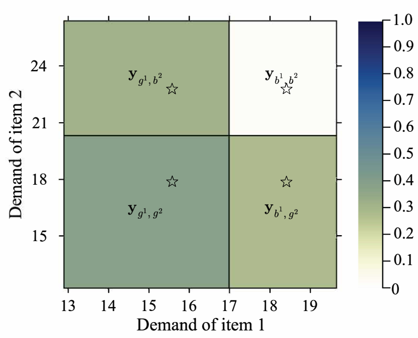

Figure 3: $\epsilon = 0.0$ (Expectation).

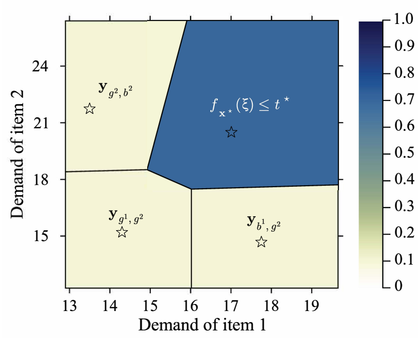

Figure 4: $\epsilon = 0.7$.

References

[1] H. Rahimian, S. Mehrotra, Frameworks and results in distributionally robust optimization, Open Journal of Mathematical Optimization 3 (2022) 1–85.

[2] D. Kuhn, S. Shafiee, W. Wiesemann, Distributionally robust optimization, Acta Numerica 34 (2025) 579–804.

[3] A. Ben-Tal, L. El Ghaoui, A. Nemirovski, Robust Optimization, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2009.

[4] A. Ardestani-Jaafari, E. Delage, Robust optimization of sums of piecewise linear functions with application to inventory problems, Operations research 64 (2016) 474–494.

[5] G. A. Hanasusanto, V. Roitch, D. Kuhn, W. Wiesemann, A distributionally robust perspective on uncertainty quantification and chance constrained programming, Mathematical Programming 151 (2015) 35–62.

[6] K. J. Cormican, D. P. Morton, R. K. Wood, Stochastic network interdiction, Operations Research 46 (1998) 184–197.

[7] J. C. Smith, Y. Song, A survey of network interdiction models and algorithms, European Journal of Operational Research 283 (2020) 797–811.

[8] D. P. Morton, F. Pan, K. J. Saeger, Models for nuclear smuggling interdiction, IIE transactions 39 (2007) 3–14.

[9] A. Abusorrah, A. Alabdulwahab, Z. Li, M. Shahidehpour, Minimax-regret robust defensive strategy against false data injection attacks, IEEE Transactions on Smart Grid 10 (2017) 2068–2079.

[10] A. Ghafouri, W. Abbas, A. Laszka, Y. Vorobeychik, X. Koutsoukos, Optimal thresholds for anomaly-based intrusion detection in dynamical environments, in: Decision and Game Theory for Security: 7th International Conference, GameSec 2016, New York, NY, USA, November 2-4, 2016, Proceedings 7, Springer, 2016, pp. 415–434.

[11] S. Kang, M. Bansal, Distributionally risk-receptive and risk-averse network interdiction problems with general ambiguity set, Networks 81 (2023) 3–22.

[12] S. Kang, M. Bansal, Distributionally ambiguous multistage stochastic integer and disjunctive programs: Applications to sequential two-player interdiction games, Virginia Tech, Blacksburg, VA, USA, Tech. Rep. SKMB-02 (2023).

[13] Ö. Elçi, N. Noyan, K. Bülbül, Chance-constrained stochastic programming under variable reliability levels with an application to humanitarian relief network design, Computers & Operations Research 96 (2018) 91–107.

[14] A. Najjarbashi, G. J. Lim, A decomposition algorithm for the two-stage chance-constrained operating room scheduling problem, IEEE Access 8 (2020) 80160–80172.

[15] C. Wan, C. Zhao, Y. Song, Chance constrained extreme learning machine for nonparametric prediction intervals of wind power generation, IEEE Transactions on Power Systems 35 (2020) 3869–3884.

[16] M. A. Quddus, S. Chowdhury, M. Marufuzzaman, F. Yu, L. Bian, A two-stage chance-constrained stochastic programming model for a bio-fuel supply chain network, International Journal of Production Economics 195 (2018) 27–44.

[17] J. Luedtke, A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support, Mathematical Programming 146 (2014) 219–244.

[18] X. Liu, S. Küçükyavuz, J. Luedtke, Decomposition algorithms for two-stage chance-constrained programs, Mathematical Programming 157 (2016) 219–243.

[19] Z. Lu, B. Zeng, Two-stage distributionally robust optimization: Intuitive understanding and algorithm development from the primal perspective, arXiv preprint arXiv:2412.20708 (2024).

[20] W. Wiesemann, D. Kuhn, M. Sim, Distributionally robust convex optimization, Operations Research 62 (2014) 1358–1376.

[21] D. Bertsimas, M. Sim, M. Zhang, Adaptive distributionally robust optimization, Management Science 65 (2019) 604–618.

[22] E. Delage, Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, Operations Research 58 (2010) 595–612.

[23] D. Bertsimas, K. Natarajan, C.-P. Teo, Persistence in discrete optimization under data uncertainty, Mathematical programming 108 (2006) 251–274.

[24] K. Natarajan, M. Song, C.-P. Teo, Persistency model and its applications in choice modeling, Management Science 55 (2009) 453–469.

[25] R. M. Van Slyke, Monte carlo methods and the pert problem, Operations research 11 (1963) 839–860.

[26] R. T. Rockafellar, S. Uryasev, Conditional Value-at-Risk for general loss distributions, Journal of Banking & Finance 26 (2002) 1443–1471.

[27] A. Shapiro, On duality theory of conic linear problems, in: Semi-Infinite Programming, Springer, 2001, pp. 135–165.

[28] W. W. Rogosinski, Moments of non-negative mass, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 245 (1958) 1–27.

[29] G. P. McCormick, Computability of global solutions to factorable nonconvex programs: Part I-Convex underestimating problems, Mathematical programming 10 (1976) 147–175.

[30] R. Gao, A. Kleywegt, Distributionally robust stochastic optimization with wasserstein distance, Mathematics of Operations Research 48 (2023) 603–655.

Appendix A. Proof of Theorem 1

Proof. We construct a discrete distribution that attains the tight upper bound $V(\mathbf{y}^{[K]})$ based on an optimal solution $(\bar{u}_k, \bar{\mathbf{w}}_k, \bar{\mathbf{z}}_k)_{k=1}^K$ of problem (3). Consider a multivariate distribution $\bar{\mathbb{P}} = \sum_{k\in[K]:\bar{u}_k>0} \bar{u}_k \delta_{(\bar{\mathbf{w}}_k/\bar{u}_k, \bar{\mathbf{z}}_k/\bar{u}_k)}$ of $(\boldsymbol{\xi}, \boldsymbol{\zeta})$, where $\delta_{(\bar{\mathbf{w}}_k/\bar{u}_k, \bar{\mathbf{z}}_k/\bar{u}_k)}$ is the Dirac distribution concentrating unit mass at $(\bar{\mathbf{w}}_k/\bar{u}_k, \bar{\mathbf{z}}_k/\bar{u}_k)$. The constructed distribution $\bar{\mathbb{P}} \in \mathcal{G}$ because

- $\sum_{k=1}^K \bar{u}_k = 1$ follows the first constraint in (3b).
- $\mathbb{E}_{\bar{\mathbb{P}}}[G\boldsymbol{\xi}] = \sum_{k\in[K]:\ \bar{u}_k>0} G\bar{u}_k(\bar{\mathbf{w}}_k/\bar{u}_k) = \sum_{k=1}^K G\bar{\mathbf{w}}_k = \boldsymbol{\mu}$. The second equality holds since for k associated with $\bar{u}_k = 0$, as we have $\bar{\mathbf{w}}_k = \mathbf{0}$ due to constraints (3d). The last equality follows the second constraint in (3b).
- $\mathbb{E}_{\bar{\mathbb{P}}}[\boldsymbol{\zeta}] = \sum_{k\in[K]:\ \bar{u}_k>0} \bar{u}_k(\bar{\mathbf{z}}_k/\bar{u}_k) \le \sum_{i=1}^K \bar{\mathbf{z}}_k \le \boldsymbol{\sigma}$. The first inequality holds due to the nonnegativity constraints (3e) on $\bar{\mathbf{z}}_k$, and the second inequality follows the last constraint in (3b).
- Since for all $k \in [K]$ associated with positive $\bar{u}_k > 0$, we have

$$O\bar{\mathbf{w}}_k + P\bar{\mathbf{z}}_k + S\,\bar{\mathbf{v}}_k \preccurlyeq_{\mathcal{K}} \mathbf{s}\bar{u}_k \Leftrightarrow O\frac{\bar{\mathbf{w}}_k}{\bar{u}_k} + P\frac{\bar{\mathbf{z}}_k}{\bar{u}_k} + S\,\frac{\bar{\mathbf{v}}_k}{\bar{u}_k} \preccurlyeq_{\mathcal{K}} \mathbf{s} \Leftrightarrow g(\frac{\bar{\mathbf{w}}_k}{\bar{u}_k}) \le \frac{\bar{\mathbf{z}}_k}{\bar{u}_k},$$

  we have $(\bar{\mathbf{w}}_k/\bar{u}_k, \bar{\mathbf{z}}_k/\bar{u}_k) \in \bar{\mathcal{U}}$.

Furthermore, the objective value of $(\bar{u}_k, \bar{\mathbf{w}}_k, \bar{\mathbf{z}}_k)_{k=1}^K$ is given as

$$\begin{aligned} V(\mathbf{y}^{[K]}) &= \sum_{k\in[K]} (Q\bar{\mathbf{w}}_k + \mathbf{q}_0\bar{u}_k)^\top \mathbf{y}^k \\ &= \sum_{i\in[K]:\bar{u}_k>0} \bar{u}_k \left[(Q\frac{\bar{\mathbf{w}}_k}{\bar{u}_i} + \mathbf{q}_0)^\top \mathbf{y}^k \right] \\ &\le \mathbb{E}_{\bar{\mathbb{P}}}\left[f(\boldsymbol{\xi})\right] \le v_0. \end{aligned}$$

The first inequality holds as we simply pick $\mathbf{y}^k$ for $(\bar{\mathbf{w}}_k/\bar{u}_k, \bar{\mathbf{z}}_k/\bar{u}_k)$ instead of solving an optimal $\mathbf{y}$-solution for $f(\bar{\mathbf{w}}_k/\bar{u}_k)$, for all $k \in [K]$ such that $\bar{u}_k > 0$. The second inequality follows from the observation $\bar{\mathbb{P}} \in \mathcal{G}$. □

Appendix B. Proof of Theorem 2

Proof. Recall that $v_0 := \sup_{\mathbb{P}\in\mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{\xi})\right]$ and $\mathcal{G}$ has $I_1 + I_2 + 1$ moment constraints of probability distributions. Following Lemma 3.1 in Shapiro [27] (also see Rogosinski [28], Theorem 1), there exists a worst-case probability distribution with a finite support of at most $I_1 + I_2 + 2$ points. Denote these points as $\{(\boldsymbol{\xi}^k, \boldsymbol{\zeta}^k)\}_{k=1}^{I_1+I_2+2}$. For each of the $I_1 + I_2 + 2$ points, calculate the optimal recourse $f(\boldsymbol{\xi}^k)$ and record $\mathbf{y}^k$ which achieves the optimal recourse. Denote $\mathcal{Y}'$ the set of distinct $\mathbf{y}$-solutions obtained and $K_0 = |\mathcal{Y}'|$. Following a similar argument in the proof of Theorem 1, we can show that $V(\mathbf{y}^{[K_0]}) = v_0$. As $\mathcal{Y}'$ is a subset of $\mathcal{Y}$, we also have $K_0 \le K$.

□

Appendix C. A Column-Constraint Generation Method

This section presents a column-constraint generation (CCG) method that starts with a relaxation containing a small subset of $\mathbf{y}$-solutions and iteratively identifies the remaining critical $\mathbf{y}$-solutions to strengthen the relaxation. We first impose the following technical assumption on the existence of a Slater point of $\bar{\mathcal{U}}$.

Assumption 2 (Slater's Condition)**.** *There exists $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\zeta}}$, $\hat{\mathbf{v}}$ such that*

$$G\hat{\boldsymbol{\xi}} = \boldsymbol{\mu},\ \hat{\boldsymbol{\zeta}} < \boldsymbol{\sigma},\ O\hat{\boldsymbol{\xi}} + P\hat{\boldsymbol{\zeta}} + S\,\hat{\mathbf{v}} \prec_{\mathcal{K}} \mathbf{s}.$$

Introducing dual variables $\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma} \geq 0, \boldsymbol{\delta}_k \succeq_{\mathcal{K}} 0, \overline{\boldsymbol{\theta}}_k \geq 0, \underline{\boldsymbol{\theta}}_k \leq 0$ with constraints (3b)–(3d), the dual of problem (3) is formulated as

$$
\begin{aligned}
V(\mathbf{y}^{[K]}) = \min_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\overline{\boldsymbol{\theta}},\underline{\boldsymbol{\theta}}} \quad & \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma} && \text{(C.1a)}\\
\text{s.t.} \quad & \alpha - \mathbf{s}^\top \boldsymbol{\delta}_k - \overline{\boldsymbol{\xi}}^\top \overline{\boldsymbol{\theta}}_k - \underline{\boldsymbol{\xi}}^\top \underline{\boldsymbol{\theta}}_k \geq \mathbf{q}_0^\top \mathbf{y}^k, \ k \in [K] && \text{(C.1b)}\\
& G^\top \boldsymbol{\beta} + O^\top \boldsymbol{\delta}_k + \overline{\boldsymbol{\theta}}_k + \underline{\boldsymbol{\theta}}_k = Q^\top \mathbf{y}^k, \ k \in [K] && \text{(C.1c)}\\
& \boldsymbol{\gamma} + P^\top \boldsymbol{\delta}^k \geq 0, \ k \in [K] && \text{(C.1d)}\\
& S^\top \boldsymbol{\delta}^k = 0, \ k \in [K] && \text{(C.1e)}\\
& \boldsymbol{\gamma} \geq 0, \ \boldsymbol{\delta}_k \succeq_{\mathcal{K}} 0, \ \overline{\boldsymbol{\theta}}_k \geq 0, \ \underline{\boldsymbol{\theta}}_k \leq 0, \ k \in [K]. && \text{(C.1f)}
\end{aligned}
$$

Under Assumption 2, strong duality holds between the primal and the dual.

Assuming only a subset $J \subset [K]$ of $\mathbf{y}$-solutions are available initially, a relaxed master problem of (3) is given as $V(\mathbf{y}^J)$. Given an optimal solution $(\hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$, the following problem is solved to identify $\mathbf{y}$-solution that is not included in set J:

$$
\begin{aligned}
\max_{\mathbf{y}\in\mathcal{Y}} \mathbf{q}_0^\top \mathbf{y} + \min_{\pi,\boldsymbol{\delta},\eta,\overline{\boldsymbol{\theta}},\underline{\boldsymbol{\theta}}} \quad & \mathbf{s}^\top \boldsymbol{\delta} + \overline{\boldsymbol{\xi}}^\top \overline{\boldsymbol{\theta}} + \underline{\boldsymbol{\xi}}^\top \underline{\boldsymbol{\theta}} && \text{(C.2a)}\\
\text{s.t.} \quad & G^\top \hat{\boldsymbol{\beta}} + O^\top \boldsymbol{\delta} + \overline{\boldsymbol{\theta}} + \underline{\boldsymbol{\theta}} = Q^\top \mathbf{y} && \text{(C.2b)}\\
& \hat{\boldsymbol{\gamma}} + P^\top \boldsymbol{\delta} \geq 0 && \text{(C.2c)}\\
& S^\top \boldsymbol{\delta} = 0 && \text{(C.2d)}\\
& \boldsymbol{\delta} \succeq_{\mathcal{K}} 0, \overline{\boldsymbol{\theta}} \geq 0, \underline{\boldsymbol{\theta}} \leq 0. && \text{(C.2e)}
\end{aligned}
$$

Let $\hat{v}$ denote the optimal value of problem (C.2) and $\hat{\mathbf{y}}$ denote an optimal solution. If $\hat{\alpha} \geq \hat{v}$, the relaxed master problem returns the true optimal value of the original worst-case bound (3). Otherwise, include $\hat{\mathbf{y}}$ in to set J and resolve the relaxed master problem.

Introducing dual variables $\tilde{\mathbf{w}}, \tilde{\mathbf{z}}, \tilde{\mathbf{v}}$ associated with with constraints (C.2b)-(C.2d), the subproblem is equivalent to

$$
\begin{aligned}
\max_{y\in\mathcal{Y},\tilde{\mathbf{w}},\tilde{\mathbf{z}},\tilde{\mathbf{v}}} \quad & \mathbf{q}_0^\top \mathbf{y} + \tilde{\mathbf{w}}^\top Q^\top \mathbf{y} - \tilde{\mathbf{w}}^\top G^\top \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\gamma}} \tilde{\mathbf{z}} && \text{(C.3a)}\\
\text{s.t.} \quad & O\tilde{\mathbf{w}} + P\tilde{\mathbf{z}} + S\tilde{\mathbf{v}} \preccurlyeq_{\mathcal{K}} \mathbf{s} && \text{(C.3b)}\\
& \underline{\boldsymbol{\xi}} \leq \tilde{\mathbf{w}} \leq \overline{\boldsymbol{\xi}} && \text{(C.3c)}\\
& \tilde{\mathbf{z}} \geq 0. && \text{(C.3d)}
\end{aligned}
$$

The resulting formulation involves bilinear terms $\tilde{w}_i y_j$ for $i = 1, \ldots, I_1, \ j = 1, \ldots, m$ in the objective function. While modern optimization solvers such as Gurobi can handle bilinear problems efficiently, these terms can also be linearized exactly using McCormick inequalities [29] as the bilinear terms can be presented as sums of products between bounded continuous variables $\tilde{w}_i$ and binary variables.

Appendix D. Proof of Theorem 3

Proof. It is clear that the formulation (4) constitutes a valid relaxation of $\sup_{\mathbb{P}\in\mathcal{G}} \mathbb{P}(f(\boldsymbol{\xi}) \geq \tau)$. To show the bound is tight, we construct a worst-case distribution that achieves the bound.

We construct a discrete distribution that attains the tight upper bound $U(\mathbf{y}^{[K]})$. Let $(\bar{u}_k, \bar{\mathbf{w}}_k, \bar{\mathbf{z}}_k, \bar{\mathbf{v}}_k)_{k=0}^K$ be an optimal solution to problem (4). Consider a multivariate distribution $\bar{\mathbb{P}} = \sum_{k\in[K]\cup\{0\}:\bar{u}_k>0} \bar{u}_k \delta_{(\bar{\mathbf{w}}_k/\bar{u}_k, \bar{\mathbf{z}}_k/\bar{u}_k)}$ of $(\boldsymbol{\xi}, \boldsymbol{\zeta})$, which represents the empirical distribution. Following a similar argument in the proof of Theorem 1, the constructed distribution $\bar{\mathbb{P}} \in \mathcal{G}$. Furthermore, the objective

$$
U(\mathbf{y}^{[K]}) = \sum_{k\in[K]} \bar{u}_k \leq \bar{\mathbb{P}}(f(\boldsymbol{\xi}) \geq \tau) \leq \sup_{\mathbb{P}\in\mathcal{G}} \mathbb{P}(f(\boldsymbol{\xi}) \geq \tau).
$$

The first inequality follows constraints (4b), i.e., for all $k \in [K]$,

$$
\begin{aligned}
& (Q\bar{\mathbf{w}}_k + q_0\bar{u}_k)^\top \mathbf{y}^k \geq \tau\bar{u}_k \\
\Leftrightarrow \quad & (Q\frac{\bar{\mathbf{w}}_k}{\bar{u}_k} + q_0)^\top \mathbf{y}^k \geq \tau, \text{ for } \bar{u}_k > 0 \\
\Rightarrow \quad & f(\frac{\bar{\mathbf{w}}_k}{\bar{u}_k}) \geq \tau, \text{ for } \bar{u}_k > 0.
\end{aligned}
$$

The second inequality follows that $\bar{\mathbb{P}} \in \mathcal{G}$.

□

Appendix E. Wasserstein Ambiguity Sets

An important class of ambiguity sets that cannot be modeled as a generalization of the SOC or the general conic ambiguity sets is Wasserstein-distance-based ambiguity sets. Consider N historical samples $\Omega := \left\{\hat{\boldsymbol{\xi}}^{\omega},\ \omega = 1, \dots, N\right\} \subset \mathcal{U}$. Let $\mathbb{P}_0 := 1/N \sum_{ß=1}^{N} \delta_{\hat{\boldsymbol{\xi}}^{\omega}}$ be the empirical distribution, where $\delta_{\hat{\boldsymbol{\xi}}^{\omega}}$ is the Dirac distribution concentrating unit mass at $\hat{\boldsymbol{\xi}}^{\omega}$. Given a positive radius ε, the Wasserstein ambiguity set is given by

$$\mathcal{W} := \left\{\mathbb{P} \in \mathcal{M}(\mathbb{R}^{I_1}) : \ \mathbb{P}(\xi \in \mathcal{U}) = 1,\ d_W(\mathbb{P}, \mathbb{P}_0) \leq \varepsilon\right\},$$

where the Wasserstein distance is defined as

$$d_W(\mathbb{P}_1, \mathbb{P}_2) := \inf_{\mathbb{Q} \sim (\mathbb{P}_1, \mathbb{P}_2)} \mathbb{E}_{\mathbb{Q}} \left(\|\boldsymbol{\xi}^1 - \boldsymbol{\xi}^2\|\right). \tag{E.1}$$

Here, $\boldsymbol{\xi}^1, \boldsymbol{\xi}^2$ are two random variables following distributions $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively, and $\mathbb{Q}$ is the coupling of $\mathbb{P}_1$ and $\mathbb{P}_2$. Denote $\mathbb{P}_\omega$ as the conditional distribution of $\mathbb{Q}$ given that $\boldsymbol{\xi}' = \hat{\boldsymbol{\xi}}^{\omega}$, $\omega \in \Omega$. The Wasserstein ambiguity set $\mathcal{W}$ is equivalent to

$$\mathcal{W} = \left\{\mathbb{P} = \frac{1}{N}\sum_{\omega \in \Omega} \mathbb{P}_\omega : \ \mathbb{P}_\omega \in \mathcal{M}(\mathbb{R}^{I_1}),\ \mathbb{P}_\omega(\boldsymbol{\xi} \in \mathcal{U}) = 1,\ \omega \in \Omega,\ \frac{1}{N}\sum_{\omega \in \Omega} \mathbb{E}_{\mathbb{P}_\omega}[\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^{\omega}\|] \leq \varepsilon\right\}.$$

Similarly, we introduce auxiliary variables $\zeta \in \mathbb{R}$ for all $\omega \in \Omega$, consider the lifted ambiguity set

$$\mathcal{V} := \left\{\mathbb{P} = \frac{1}{N}\sum_{\omega \in \Omega} \mathbb{P}_\omega : \ \mathbb{P}_\omega \in \mathcal{M}(\mathbb{R}^{I_1+1}),\ \mathbb{P}_\omega((\boldsymbol{\xi}, \zeta) \in \bar{\mathcal{U}}_\omega) = 1,\ \omega \in \Omega,\ \frac{1}{N}\sum_{\omega \in \Omega} \mathbb{E}_{\mathbb{P}_\omega}[\zeta] \leq \varepsilon\right\}.$$

For all $\omega \in \Omega$, the lifted ambiguity set $\bar{\mathcal{U}}_\omega$ is defined as

$$\bar{\mathcal{U}}_\omega := \left\{(\boldsymbol{\xi}, \zeta) \in \mathbb{R}^{I_1+1} : \ \boldsymbol{\xi} \in \mathcal{U},\ \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^{\omega}\| \leq \zeta\right\}.$$

We assume that the support set $\mathcal{U}$ can be represented using a proper cone $\mathcal{K}$ as $\mathcal{U} = \left\{\boldsymbol{\xi} \in \mathbb{R}^{I_1} : \ \exists \mathbf{v} \in \mathbb{R}^{I_3}, O\boldsymbol{\xi} + S\,\mathbf{v} \preccurlyeq_{\mathcal{K}} \mathbf{s}\right\}$.

Proposition 1. *The Wasserstein ambiguity set $\mathcal{W}$ is equivalent to the set of marginal distributions of $\boldsymbol{\xi}$ under all distributions in the lifted set $\mathcal{V}$.*

Proof. The proof is similar to Wiesemann et al. [20] and Bertsimas et al. [21]. We first show that the marginal distributions of $\boldsymbol{\xi}$ under any $\mathbb{P} \in \mathcal{V}$ belongs to set $\mathcal{W}$. For any $\mathbb{P} = 1/N \sum_{\omega \in \Omega} \mathbb{P}_\omega \in \mathcal{V}$, let $\mathbb{P}^*$ be the marginal distribution of $\boldsymbol{\xi}$ under $\mathbb{P}$ and $\mathbb{P}^*_\omega$ be the marginal distribution of $\boldsymbol{\xi}$ under $\mathbb{P}_\omega$ for all $\omega \in \Omega$.

1. For all $\omega \in \Omega$, since $\mathbb{P}_\omega((\boldsymbol{\xi}, \zeta) \in \bar{\mathcal{U}}_\omega) = 1$, we have $\mathbb{P}_\omega(\boldsymbol{\xi} \in \mathcal{U}) = 1$ and $\mathbb{P}_\omega(\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^{\omega}\| \leq \zeta) = 1$. It follows that $\mathbb{P}^*_\omega(\boldsymbol{\xi} \in \mathcal{U}) = 1$.
2. Since $1/N \sum_{\omega \in \Omega} \mathbb{E}_{\mathbb{P}_\omega}[\zeta] \leq \varepsilon$ and $\mathbb{E}_{\mathbb{P}^*_\omega}[\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^{\omega}\|] = \mathbb{E}_{\mathbb{P}_\omega}[\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^{\omega}\|] \leq \mathbb{E}_{\mathbb{P}_\omega}[\zeta]$, for all $\omega \in \Omega$, we have $1/N \sum_{\omega \in \Omega} \mathbb{E}_{\mathbb{P}^*_\omega}[\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^{\omega}\|] \leq \varepsilon$.

Hence, $\mathbb{P}^* = 1/N \sum_{\omega\Omega} \mathbb{P}^*_\omega \in \mathcal{W}$.

Conversely, consider $\mathbb{P}^* = 1/N \sum_{\omega\in\Omega} \mathbb{P}^*_\omega \in \mathcal{W}$ and observe that $\mathbb{P}^*_\omega\left((\boldsymbol{\xi}, \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^\omega\|) \in \mathcal{U}_\omega\right) = 1,\ \omega \in \Omega$. Hence, one can construct a probability distribution $\mathbb{P}_\omega \in \mathcal{M}(\mathbb{R}^{I_1+1})$ for random variable $(\boldsymbol{\xi}, \zeta) \in \bar{\mathcal{U}}_\omega$ such that $(\boldsymbol{\xi}, \zeta) = (\boldsymbol{\xi}, \boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^\omega)$ almost surely under $\mathbb{P}_\omega$. Thus we have $\mathbb{P}_\omega((\boldsymbol{\xi}, \zeta) \in \mathcal{U}_\omega) = 1$ and $\mathbb{E}_{\mathbb{P}_\omega}[\zeta] = \mathbb{E}_{\mathbb{P}_\omega}[\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^\omega\|] = \mathbb{E}_{\mathbb{P}^*_\omega}[\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^\omega\|]$. It follows that $1/N \sum_{\omega\in\Omega} \mathbb{E}_{\mathbb{P}_\omega}[\zeta] = 1/N \sum_{\omega\in\Omega} \mathbb{E}_{\mathbb{P}^*_\omega}[\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^\omega\|] \le \varepsilon$. Therefore $\mathbb{P} \in \mathcal{V}$.

□

The convex representation based on all feasible **y**-solutions also applies under the lifted Wasserstein ambiguity set $\mathcal{V}$. The key difference is that $N = |\Omega|$ copies of u-, **w**-, and z-variables are created, each for a conditional $\mathbb{P}_\omega$ associated with $\hat{\boldsymbol{\xi}}^\omega$.

Corollary 3. *Given all the feasible solutions of $\mathcal{Y} = \{\mathbf{y}^k\}_{k=1}^K$, the worst-case expectation $\sup_{\mathbb{P}\in\mathcal{W}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{\xi})]$ equals to*

$$V_{\mathcal{V}}(\mathbf{y}^{[K]}) := \max_{(u_k^\omega, \mathbf{w}_k^\omega, z_k^\omega, \mathbf{v}_k^\omega, \omega\in\Omega)_{k=1}^K} \quad \frac{1}{N} \sum_{\omega\in\Omega} \sum_{k=1}^K (Q\mathbf{w}_k^\omega + \mathbf{q}_0 u_k^\omega)^\top \mathbf{y}^k \tag{E.2a}$$

$$s.t. \quad \sum_{k=1}^K u_k^\omega = 1,\ \omega \in \Omega \tag{E.2b}$$

$$\frac{1}{N} \sum_{\omega\in\Omega} \sum_{k=1}^K z_k^\omega \le \varepsilon \tag{E.2c}$$

$$O\mathbf{w}_k^\omega + S\mathbf{v}_k^\omega \preccurlyeq_{\mathcal{K}} \mathbf{s}u_k^\omega,\ k \in [K],\ \omega \in \Omega \tag{E.2d}$$

$$\|\mathbf{w}_k^\omega - u_k^\omega \hat{\boldsymbol{\xi}}^\omega\| \le z_k^\omega,\ k \in [K],\ \omega \in \Omega \tag{E.2e}$$

$$\bar{\boldsymbol{\xi}} u_k^\omega \ge \mathbf{w}_k^\omega \ge \underline{\boldsymbol{\xi}} u_k^\omega,\ k \in [K],\ \omega \in \Omega \tag{E.2f}$$

$$u_k^\omega \ge \mathbf{0},\ z_k^\omega \ge \mathbf{0},\ k \in [K],\ \omega \in \Omega. \tag{E.2g}$$

The proof is similar to that of Theorem 2 and thus omitted. We can also establish a similar finite convex formulation result as in Theorem 2.

Corollary 4. *There exists a subset $\mathcal{Y}' \subset \mathcal{Y}$ of size $K_0 \le \min\{N+1, K\}$ such that $V_{\mathcal{V}}(\mathbf{y}^{[K_0]}) = \sup_{\mathbb{P}\in\mathcal{W}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{\xi})]$.*

Proof. Recall that $\mathcal{W}$ has N-point-supported nominal distribution $\mathbb{P}_0$. Following Corollary 2(ii) in Gao and Kleywegt [30], there exists a worst-case distribution supported on at most $N+1$ points. Denote them as $\boldsymbol{\xi}^k,\ k = 1, \ldots, N+1$. For each of them, calculate the optimal recourse $f(\boldsymbol{\xi}^k)$ and record their corresponding optimal solution $\mathbf{y}^k$. note $\mathcal{Y}'$ the set of distinct **y**-solutions obtained and $K_0 = |\mathcal{Y}'|$. Following a similar argument in the proof of Corollary 3, we can show that $V_{\mathcal{V}}(\mathbf{y}^{[K_0]}) = \sup_{\mathbb{P}\in\mathcal{W}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{\xi})]$. As $\mathcal{Y}'$ is a subset of $\mathcal{Y}$, we also have $K_0 \le K$. □