

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

Yiling Zhang^a, Zhe Li^b, Eojin Han^c

^aIndustrial and Systems Engineering, University of Minnesota, Minneapolis, 55455, MN, USA

^bChemical Engineering and Materials Science, University of Minnesota, Minneapolis, 55455, MN, USA

^cMendoza College of Business, University of Notre Dame, Notre Dame, 46556, IN, USA

Abstract

This paper studies distributionally robust optimization problems with integer recourse. We develop a unified framework that yields exact finite convex reformulations under both conic moment-based and Wasserstein ambiguity sets. The resulting formulations admit tractable primal representations without relying on sampling or semi-infinite optimization techniques. Beyond tractability, the proposed framework provides an explicit characterization of worst-case distributions in terms of critical recourse decisions. This characterization offers new interpretability, generalizes persistence analysis, and provides a theoretical foundation for column-and-constraint generation algorithms. The framework applies to a broad class of risk measures, including expectation, probability, and conditional value-at-risk (CVaR). In particular, it enables finite reformulations and decomposition algorithms for two-stage distributionally robust chance-constrained problems with integer recourse, a class of problems that has received limited attention in the literature. Numerical experiments demonstrate the computational tractability of the proposed approach and the effectiveness of the resulting algorithms.

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

1. Introduction

We consider an optimization problem in the following form:

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

where $f(\xi) := \max_{\mathbf{y} \in \mathcal{Y}} \mathbf{q}(\xi)^\top \mathbf{y}$ and $\mathcal{Y} \subset \mathbb{R}^m$ is either a discrete set or a polyhedron. The objective coefficient $\mathbf{q}(\xi) : \mathbb{R}^{I_1} \mapsto \mathbb{R}^m$ is assumed linear in the random variable $\xi \in \mathbb{R}^{I_1}$. Specifically, $\mathbf{q}(\xi) = Q\xi + \mathbf{q}_0$ with $Q \in \mathbb{R}^{m \times I_1}$ 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 ξ . Here, the ambiguity set $\mathcal{D}$ captures partial information on the unknown underlying true probability distribution. For simplicity, we assume that the set of ξ for which $f(\xi)$ admits multiple optimal solutions has 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) [20, 13], 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:

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

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 $\mathbf{q}(\xi) = \mathbf{h}(\xi) + T(\xi)\mathbf{x}$, where both $\mathbf{h}(\xi)$ and $T(\xi)$ are linear in ξ . Here, we use $f_{\mathbf{x}}(\xi)$ 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 involve the integer recourse structure of distributionally robust bounds (1), particularly in the context of the minimax formulation (2). Below, we first describe three representative examples to illustrate the generality of this setting and then discuss the methodology relevance and outline the main contributions.

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 $\boldsymbol{\xi}$: $f_{\mathbf{x}}(\boldsymbol{\xi}) = \sum_{\ell=1}^L \max_{k=1, \dots, K_{\ell}} c_{\ell k}^0(\boldsymbol{\xi}) + \mathbf{c}_{\ell k}(\boldsymbol{\xi})^{\top} \mathbf{x}$, which is widely used in multi-item newsvendor problems and related inventory problems [see, e.g., 2, 10]. This can be equivalently written as

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

where $\mathcal{Y} = \{(\mathbf{y}_1, \dots, \mathbf{y}_L) : \mathbf{y}_{\ell} \in \{0,1\}^{K_{\ell}}, \mathbf{e}^{\top} \mathbf{y}_{\ell} = 1 \ \forall \ell = 1, \dots, L\}$. Here, $\mathbf{c}_{\ell}^0(\boldsymbol{\xi}) := [c_{\ell k}^0(\boldsymbol{\xi})]_{k \in [K_{\ell}]}$, $C_{\ell}(\boldsymbol{\xi})$ 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 $\mathbf{c}_{\ell}^0(\boldsymbol{\xi})$, $C_{\ell}(\boldsymbol{\xi})$, and $\mathbf{y}_{\ell}$, respectively.

Risk-averse network interdiction problems. Network interdiction problems [6, 23] 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 [16], and ensuring resilience and security in critical infrastructure [1, 9]. The risk-averse and DRO formulations [12, 11] 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 [8], healthcare operations [17], power systems [25], and supply chain [19]. 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 [15, 14] is written as

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

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. The proposed framework builds on several streams of DRO, including moment-based and Wasserstein ambiguity sets, finite-support worst-case distributions, and decomposition-based solution methods. Our work makes three main contributions.

First, we develop a unified framework for distributionally robust bounds with integer recourse that yields finite convex reformulations for a broad class of ambiguity sets, including moment-based and Wasserstein sets, and for several risk measures, including expectation, probability, and conditional value-at-risk (CVaR). The framework provides a systematic treatment across different ambiguity sets and risk measures.

Second, the reformulations are derived from a primal perspective rather than the duality-based semi-infinite programming approaches commonly used in the DRO literature [13]. This perspective establishes an explicit connection between worst-case distributions and the critical recourse decisions they induce, providing structural insight into the underlying optimization problem. These insights naturally lead to the development of the column-and-constraint generation (CCG) methods for DRO with integer recourse, while also explaining the effectiveness of these approaches.

Third, building on this characterization of worst-case distributions and critical recourse decisions, we generalize decision persistence [4, 18], which quantifies not only the likelihood of selecting a recourse decision, but also its contribution to the worst-case objective and the variability of the uncertainty realizations under which it is selected. The framework further extends persistence analysis to richer ambiguity descriptions, including partial cross-moment information and structured support sets. Finally, we demonstrate the applicability of the proposed framework through illustrative examples and computational experiments.

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 detailed proofs, column-and-constraint generation algorithms, extensions to Wasserstein ambiguity sets and CVaR-based formulations, and additional computational experiments.

Notation. We use the shorthand $[N] = \{1, \dots, N\}$ 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., 26, 5]. Our results can be extended to ambiguity sets based on generalized cones and Wasserstein metrics; the latter is presented in Appendix D.

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

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

where $G \in \mathbb{R}^{L_1 \times I_1}$, $\mu \in \mathbb{R}^{L_1}$, and $\sigma \in \mathbb{R}^{I_2}$. The last constraint is a componentwise inequality, i.e., $\mathbb{E}_{\mathbb{P}}[g_i(\xi)] \leq \sigma_i$, $i \in [I_2]$. 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., $\text{epi}(g_i) := \{(\xi, z) \in \mathbb{R}^{I_1+1} : g_i(\xi) \leq z\}$ 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. [5] for more distributional characterizations.

Following [26, 5], we introduce a lifting variable $\zeta \in \mathbb{R}^{I_2}$ and consider the following *lifted* ambiguity set

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

where $\bar{\mathcal{U}}$ is the lifted support set defined as $\bar{\mathcal{U}} := \{(\xi, \zeta) \in \mathbb{R}^{I_1+I_2} : \xi \in \mathcal{U}, g(\xi) \leq \zeta\}$, with $g(\xi) = (g_1(\xi), \dots, g_{I_2}(\xi))$. Given the SOC representable support set $\mathcal{U}$ and epigraphs $\text{epi}(g_i)$, $i \in [I_2]$, 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 $\bar{\mathcal{U}} = \{(\xi, \zeta) \in \mathbb{R}^{I_1+I_2} : \exists \mathbf{v} \in \mathbb{R}^{I_3}, O\xi + P\zeta + S\mathbf{v} \preceq_{\mathcal{K}} \mathbf{s}\}$. An important observation is that the ambiguity set $\mathcal{D}$ is equivalent to the set of marginal distributions of ξ under all distributions in the lifted set $\mathcal{G}$ [see Proposition 1, 5], 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., $\sup_{\mathbb{P} \in \mathcal{D}} \rho_{\mathbb{P}}[f(\xi)] = \sup_{\mathbb{P} \in \mathcal{G}} \rho_{\mathbb{P}}[f(\xi)]$.*

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

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

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 [7] 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 $\xi \in \mathcal{U}$, let $\hat{\mathbf{y}}(\xi)$ denote the optimal solution of the recourse function $f(\xi)$. Since ξ is random, $\hat{\mathbf{y}}(\xi)$ is a random variable depending on ξ . Given a distribution $\mathbb{P} \in \mathcal{G}$, for $k = 1, \dots, K$, let

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

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 $\tilde{\mathcal{U}}$, we have $O\mathbf{w}_k + P\mathbf{z}_k + S\mathbf{v}_k \preceq_{\mathcal{H}} \mathbf{s}u_k$ and $\bar{\xi}u_k \geq \mathbf{w}_k \geq \underline{\xi}u_k$ for some $\mathbf{v}_k$, $k \in [K]$.

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

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

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 $f(\boldsymbol{\xi})$ in (1) equals to*

$$V(\mathbf{y}^{[K]}) := \max_{(u_k, \mathbf{w}_k, \mathbf{z}_k, \mathbf{v}_k)_{k=1}^K} \sum_{k=1}^K (Q\mathbf{w}_k + \mathbf{q}_0 u_k)^\top \mathbf{y}^k \quad (3a)$$

$$s.t. \quad \sum_{k=1}^K u_k = 1, \quad G \sum_{k=1}^K \mathbf{w}_k = \boldsymbol{\mu}, \quad \sum_{k=1}^K \mathbf{z}_k \leq \boldsymbol{\sigma} \quad (3b)$$

$$O\mathbf{w}_k + P\mathbf{z}_k + S\mathbf{v}_k \preceq_{\mathcal{H}} \mathbf{s}u_k, \quad k \in [K] \quad (3c)$$

$$\bar{\xi}u_k \geq \mathbf{w}_k \geq \underline{\xi}u_k, \quad k \in [K] \quad (3d)$$

$$u_k \geq 0, \quad \mathbf{z}_k \geq 0, \quad k \in [K]. \quad (3e)$$

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. Specifically, the constructed worst-case distribution admits a recourse-solution-based characterization that explicitly associates support points of the distribution to recourse-decisions. Notably, the solution of the formulation above quantifies 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 [4, 18], 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 $\mathbf{v}_k$.

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). Theorem 2 suggests that a relatively small number of recourse decisions is sufficient to exactly solve problem (1). In practice, the number of feasible recourse $\mathbf{y}$ can be much smaller depending on the distribution of $\mathbf{q}(\boldsymbol{\xi})$. Intuitively, if the distribution of $\mathbf{q}(\boldsymbol{\xi})$ is more concentrated around its nominal value, a subset of very few recourse actions may become active with strictly positive probabilities.

Remark (Recourse Solution Enumeration and Column-and-Constraint Generation). 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 explicitly, one can instead start with a small subset of recourse solutions and iteratively identify additional ones as needed. To achieve this, we develop a column-constraint generation method with the details provided in Appendix C. Theorem 2 provides the theoretical foundation for this approach by establishing that a tight relaxation requires only a limited number of recourse solutions, bounded by $I_1 + I_2 + 2$. Consequently, the proposed CCG method terminates after generating a small subset of recourse solutions.

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 D. Additionally, since CVaR admits a convex optimization reformulation involving its expectation [22], the finite convex reformulation extends directly to CVaR-based DRO problems. The details are presented in Appendix E.

3. Probability Bound

In this section, we consider a probability measure in (1): $\rho_{\mathbb{P}}(f(\xi)) = \mathbb{P}(f(\xi) \geq \tau)$, 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 $\xi \in \mathcal{U}$, let $\hat{\mathbf{y}}(\xi)$ denote the optimal solution of the recourse function $f(\xi)$ such that $f(\xi) \geq \tau$. Since ξ is random, $\hat{\mathbf{y}}(\xi)$ is a random variable depending on ξ . Given a distribution $\mathbb{P} \in \mathcal{D}$, for $k = 1, \dots, K$, let

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

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

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

Moreover, we introduce $u_0 = \mathbb{P}(f(\xi) < \tau)$, $\mathbf{w}_0 = \mathbb{E}_{\mathbb{P}}[\xi \mid f(\xi) < \tau] \mathbb{P}(f(\xi) < \tau)$, and $\mathbf{z}_0 = \mathbb{E}_{\mathbb{P}}[\zeta \mid f(\xi) < \tau] \mathbb{P}(f(\xi) < \tau)$. 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 $\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{P}(f(\xi) \geq \tau)$ equals to*

$$U(\mathbf{y}^{[K]}) := \max_{(u_k, \mathbf{w}_k, \mathbf{z}_k, \mathbf{v}_k)_{k=0}^K} \sum_{k=1}^K u_k \quad (4a)$$

$$s.t. \quad (Q\mathbf{w}_k + q_0 u_k)^{\top} \mathbf{y}^k \geq \tau u_k, \quad k \in [K] \quad (4b)$$

$$\sum_{k=0}^K u_k = 1, \quad G \sum_{k=0}^K \mathbf{w}_k = \boldsymbol{\mu}, \quad \sum_{k=0}^K \mathbf{z}_k \leq \boldsymbol{\sigma} \quad (4c)$$

$$O\mathbf{w}_k + P\mathbf{z}_k + S\mathbf{v}_k \preceq_{\mathcal{K}} S u_k, \quad k \in [K] \cup \{0\} \quad (4d)$$

$$\bar{\xi} u_k \geq \mathbf{w}_k \geq \underline{\xi} u_k, \quad k \in [K] \cup \{0\} \quad (4e)$$

$$\mathbf{z}_k \geq 0, \quad u_k \geq 0, \quad k \in [K] \cup \{0\}. \quad (4f)$$

The proof of Theorem 3 is presented in Appendix F. 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 $U(\mathbf{y}^{[K_0]}) = v_0$.*

Similar to Theorems 1-2, Theorem 3 and Corollary 2 establish an explicit connection between worst-case distributions and critical recourse decision. This connection generalizes persistence analysis, and motivates the development of a CCG method that only needs to generate a limited number of recourse solutions. The details of the CCG method is presented in Appendix H.1. We illustrate these results through a distributionally robust network interdiction problem and report numerical experiments in Appendix H.

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 v12.0.3. 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. [4] and Van Slyke [24], 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.

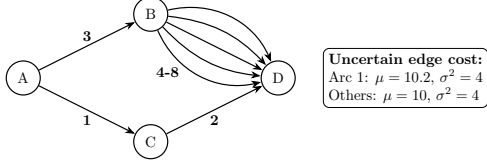


Figure 1: Project network structure.

Methods	Expected Duration	Probability (Criticality)					
		3 → 4	3 → 5	3 → 6	3 → 7	3 → 8	1 → 2
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

Table 1: Comparison of expected duration and path criticalities.

Assuming only marginal distributional information, Bertsimas et al. [4] 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 μ and standard deviation σ for all activities, with a lower bound $\xi = \mu - 3\sigma$ and an upper bound $\bar{\xi} = \mu + 3\sigma$. We compare two models: MM using only marginal moments, and PCM additionally incorporating partial crossmoments of the form

$$\mathbb{E}_{\mathbb{P}}[(\mathbf{f}_i^{\top}(\xi - \mu))^2] \leq \mathbf{f}_i^{\top} \Sigma \mathbf{f}_i, \quad i = 1, \dots, 8, \quad \text{where } \mathbf{f}_{ij} = \begin{cases} 2 & \text{if } j = i, \\ -\frac{1}{4} & \text{otherwise} \end{cases}, \quad i = 1, \dots, 8.$$

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. As a benchmark, we conduct a Monte Carlo simulation (MCS) by drawing 50,000 samples from a multivariate normal distribution with mean μ and covariance Σ . Any samples with negative values should be truncated to ensure non-negative activity durations; however, all samples were non-negative for the parameter settings in this study.

Table 1 shows the expected duration and criticality probability of each path. The PCM model yields a smaller expected completion time than MM due to the incorporation of partial crossmoment information. More importantly, while MM assigns equal criticality probabilities to all five paths passing through arc 3, PCM differentiates these paths through a ranking that better aligns with the MCS benchmark. This example illustrates that incorporating partial cross-moment information can improve the identification of critical paths. In addition, our framework extends the persistence analysis of Bertsimas et al. [4] by providing richer statistics, including objective contribution and the variability of uncertainty realizations associated with each path.

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 $g_n(x) = \min_{i=1,2,3} \alpha_n^i x_n + \beta_n^i$ is a piecewise concave function with three pieces, while the goodwill cost $b_n(x) = \max_{i=1,2,3} \alpha_n^i x_n + \beta_n^i$ 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

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

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.

For product n , the uncertain demand is modeled as $\xi_n = \psi + \omega_n$, where ψ is a systematic factor common to all products, and ω_n is an idiosyncratic factor specific to product n . The systematic factor follows a uniform distribution $\psi \sim \mathcal{U}(12, 15)$, while the idiosyncratic factor follows a Gaussian distribution $\omega_n \sim \mathcal{N}(3n, n)$. We use an ambiguity set with partial cross-moment ambiguity set with $\mathbb{E}_{\mathbb{P}}[(\mathbf{1}^{\top}(\xi - \mu))^2] \leq \mathbf{1}^{\top} \Sigma \mathbf{1}$. To fit the ambiguity set, we draw 500 samples to estimate the mean μ and covariance Σ , and pointwise minimum and maximum as the support. As the number of items n increases, directly solving the finite convex reformulation becomes computationally challenging. Thus we use the proposed CCG method (presented in Appendix C) to solve all instances. Detailed computational results are presented in Appendix G.

N	2	4	8	16	32
K_0	4	5	10	16	33
$I_1 + I_2 + 2$	7	11	19	35	67

Table 2: Number of $\mathbf{y}$ with positive probabilities.

$\hat{\mathbf{y}}$	$\mathbb{P}(\hat{\mathbf{y}})$	$\mathbb{E}_{\mathbb{P}}[f_{\mathbf{x}^*}(\xi) \hat{\mathbf{y}}]$	$\text{Var}_{\mathbb{P}}[f_{\mathbf{x}^*}(\xi) \hat{\mathbf{y}}]$
$\mathbf{y}_{g^1, g^2}$	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 3: 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, $K_0 \leq \min\{K, I_1 + I_2 + 2\}$. Table 2 shows their values for newsvendor problems using expectation with $N \in \{2, 4, 8, 16, 32\}$ 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 (see Appendix E), which measures the expected loss exceeding the value-at-risk (VaR), the ϵ -percentile [21]. The expectation measure can be viewed as a special case of CVaR when $\epsilon = 0.0$. Next, we examine how critical recourse 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 $y_4 = y_{10} = 1$. That is, both items have overage and the excessive amount triggers their first piece of the salvage functions.

Table 3 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 2-3 illustrate the partition of the support set and the associated optimal recourse solutions. In Figure 3, the dark colored region represents the realizations of the recourse function below the $\text{VaR} = -338.08$ 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.

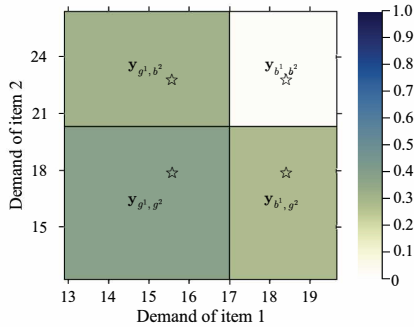


Figure 2: $\epsilon = 0.0$ (Expectation).

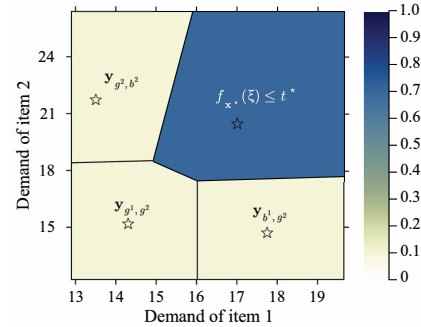


Figure 3: $\epsilon = 0.7$.

References

- [1] Abusorrah, A., Alabdulwahab, A., Li, Z., Shahidehpour, M., 2017. Minimax-regret robust defensive strategy against false data injection attacks. *IEEE Transactions on Smart Grid* 10, 2068–2079.
- [2] Ardestani-Jaafari, A., Delage, E., 2016. Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations research* 64, 474–494.
- [3] Ben-Tal, A., El Ghaoui, L., Nemirovski, A., 2009. Robust Optimization. Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ.
- [4] Bertsimas, D., Natarajan, K., Teo, C.P., 2006. Persistence in discrete optimization under data uncertainty. *Mathematical programming* 108, 251–274.
- [5] Bertsimas, D., Sim, M., Zhang, M., 2019. Adaptive distributionally robust optimization. *Management Science* 65, 604–618.

- [6] Cormican, K.J., Morton, D.P., Wood, R.K., 1998. Stochastic network interdiction. *Operations Research* 46, 184–197.
- [7] Delage, E., Ye, Y., 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 58, 595–612.
- [8] Elçi, Ö., Noyan, N., Bülbül, K., 2018. Chance-constrained stochastic programming under variable reliability levels with an application to humanitarian relief network design. *Computers & Operations Research* 96, 91–107.
- [9] Ghafouri, A., Abbas, W., Laszka, A., Vorobeychik, Y., Koutsoukos, X., 2016. 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. pp. 415–434.
- [10] Hanasusanto, G.A., Roitch, V., Kuhn, D., Wiesemann, W., 2015. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming* 151, 35–62.
- [11] Kang, S., Bansal, M., 2023a. Distributionally ambiguous multistage stochastic integer and disjunctive programs: Applications to sequential two-player interdiction games. Virginia Tech, Blacksburg, VA, USA, Tech. Rep. SKMB-02 .
- [12] Kang, S., Bansal, M., 2023b. Distributionally risk-receptive and risk-averse network interdiction problems with general ambiguity set. *Networks* 81, 3–22.
- [13] Kuhn, D., Shafiee, S., Wiesemann, W., 2025. Distributionally robust optimization. *Acta Numerica* 34, 579–804.
- [14] Liu, X., Küçükyavuz, S., Luedtke, J., 2016. Decomposition algorithms for two-stage chance-constrained programs. *Mathematical Programming* 157, 219–243.
- [15] Luedtke, J., 2014. A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Mathematical Programming* 146, 219–244.
- [16] Morton, D.P., Pan, F., Saeger, K.J., 2007. Models for nuclear smuggling interdiction. *IIE transactions* 39, 3–14.
- [17] Najjarbashi, A., Lim, G.J., 2020. A decomposition algorithm for the two-stage chance-constrained operating room scheduling problem. *IEEE Access* 8, 80160–80172.
- [18] Natarajan, K., Song, M., Teo, C.P., 2009. Persistency model and its applications in choice modeling. *Management Science* 55, 453–469.
- [19] Quddus, M.A., Chowdhury, S., Marufuzzaman, M., Yu, F., Bian, L., 2018. A two-stage chance-constrained stochastic programming model for a bio-fuel supply chain network. *International Journal of Production Economics* 195, 27–44.
- [20] Rahimian, H., Mehrotra, S., 2022. Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization* 3, 1–85.
- [21] Rockafellar, R.T., Uryasev, S., 2002. Conditional Value-at-Risk for general loss distributions. *Journal of Banking & Finance* 26, 1443–1471.
- [22] Rockafellar, R.T., Uryasev, S., et al., 2000. Optimization of conditional value-at-risk. *Journal of risk* 2, 21–42.
- [23] Smith, J.C., Song, Y., 2020. A survey of network interdiction models and algorithms. *European Journal of Operational Research* 283, 797–811.
- [24] Van Slyke, R.M., 1963. Monte carlo methods and the pert problem. *Operations research* 11, 839–860.
- [25] Wan, C., Zhao, C., Song, Y., 2020. Chance constrained extreme learning machine for nonparametric prediction intervals of wind power generation. *IEEE Transactions on Power Systems* 35, 3869–3884.
- [26] Wiesemann, W., Kuhn, D., Sim, M., 2014. Distributionally robust convex optimization. *Operations Research* 62, 1358–1376.

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 (ξ, ζ) , 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).
- $\mathbb{E}_{\bar{\mathbb{P}}}[G\xi] = \sum_{k \in [K]: \bar{u}_k > 0} G\bar{u}_k(\bar{\mathbf{w}}_k/\bar{u}_k) = \sum_{k=1}^K G\bar{\mathbf{w}}_k = \boldsymbol{\mu}$. 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).
- $\mathbb{E}_{\bar{\mathbb{P}}}[\zeta] = \sum_{k \in [K]: \bar{u}_k > 0} \bar{u}_k(\bar{\mathbf{z}}_k/\bar{u}_k) \leq \sum_{i=1}^K \bar{\mathbf{z}}_k \leq \boldsymbol{\sigma}$. 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

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

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

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

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 $\bar{\mathbb{P}} \in \mathcal{G}$. $\square$

Appendix B. Proof of Theorem 2

Proof. Recall that $v_0 := \sup_{\mathbb{P} \in \mathcal{G}} \mathbb{E}_{\mathbb{P}}[f(\xi)]$ and $\mathcal{G}$ has $I_1 + I_2 + 1$ moment constraints of probability distributions which is measurable. Following Lemma 3.1 in Shapiro [6] (also see Rogosinski [5], 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 $\{(\xi^k, \zeta^k)\}_{k=1}^{I_1+I_2+2}$. For each of the $I_1 + I_2 + 2$ points, calculate the optimal recourse $f(\xi^k)$ 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 $K_0 \leq K$. $\square$

Appendix C. Solution Strategy: Master-Subproblem Scheme

The finite convex reformulation in Theorem 1 establishes that the exact worst-case bound can be attained using only a finite number of recourse solutions. However, the set of feasible recourse solutions $\mathcal{Y}$ may still be prohibitively large to enumerate explicitly. This observation naturally motivates decomposition strategies that identifies only the recourse solutions required to attain the worst-case bound v_0 in (1).

In this section, we develop two decomposition methods. The first is a column-and-constraint generation (CCG) algorithm that iteratively adds critical recourse solutions together with their associated dual variables and structural constraints, exploiting the finite-recourse representation presented in Theorem 1. The second is a Benders-style cutting-plane algorithm that tightens the approximation of the recourse function through a sequence of supporting hyperplanes.

We first impose the following technical assumption to guarantee strong duality.

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

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

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

$$V(\mathbf{y}^{[K]}) = \min_{\alpha, \beta, \gamma, \delta, \bar{\theta}, \underline{\theta}} \alpha + \mu^\top \beta + \sigma^\top \gamma \quad (\text{C.1a})$$

$$\text{s.t.} \quad \alpha - \mathbf{s}^\top \delta_k - \bar{\xi}^\top \bar{\theta}_k - \underline{\xi}^\top \underline{\theta}_k \geq \mathbf{q}_0^\top \mathbf{y}^k, \quad k \in [K] \quad (\text{C.1b})$$

$$G^\top \beta + O^\top \delta_k + \bar{\theta}_k + \underline{\theta}_k = Q^\top \mathbf{y}^k, \quad k \in [K] \quad (\text{C.1c})$$

$$\gamma + P^\top \delta^k \geq 0, \quad k \in [K] \quad (\text{C.1d})$$

$$S^\top \delta^k = 0, \quad k \in [K] \quad (\text{C.1e})$$

$$\gamma \geq 0, \delta_k \geq_{\mathcal{K}} 0, \bar{\theta}_k \geq 0, \underline{\theta}_k \leq 0, \quad k \in [K], \quad (\text{C.1f})$$

which we refer to as the full formulation. Under Assumption 2, strong duality holds between the primal (3) and the dual (C.1). When only a subset $\mathcal{J} \subset [K]$ of $\mathbf{y}$ -solutions are available initially, a relaxed master problem is given as $V(\mathbf{y}^{\mathcal{J}})$. Given an optimal solution $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ to the relaxed master problem $V(\mathbf{y}^{\mathcal{J}})$, the following subproblem is solved to identify $\mathbf{y}$ -solution that is not included in set $\mathcal{J}$:

$$\max_{\mathbf{y} \in \mathcal{Y}} \mathbf{q}_0^\top \mathbf{y} + \min_{\pi, \delta, \eta, \bar{\theta}, \underline{\theta}} \mathbf{s}^\top \delta + \bar{\xi}^\top \bar{\theta} + \underline{\xi}^\top \underline{\theta} \quad (\text{C.2a})$$

$$\text{s.t.} \quad G^\top \hat{\beta} + O^\top \delta + \bar{\theta} + \underline{\theta} = Q^\top \mathbf{y} \quad (\text{C.2b})$$

$$\hat{\gamma} + P^\top \delta \geq 0 \quad (\text{C.2c})$$

$$S^\top \delta = 0 \quad (\text{C.2d})$$

$$\delta \geq_{\mathcal{K}} 0, \bar{\theta} \geq 0, \underline{\theta} \leq 0. \quad (\text{C.2e})$$

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 optimal solution $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ to the relaxed master problem is optimal to the original worst-case bound formulation in (1). Otherwise, the recourse solution $\hat{\mathbf{y}}$ identifies a violated constraint of the full formulation (C.1). Therefore, the recourse $\hat{\mathbf{y}}$ is added to the set $\mathcal{J}$ and we resolve the relaxed master problem.

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

$$\max_{\mathbf{y} \in \mathcal{Y}, \tilde{\mathbf{w}}, \tilde{\mathbf{z}}, \tilde{\mathbf{v}}} \mathbf{q}_0^\top \mathbf{y} + \tilde{\mathbf{w}}^\top Q^\top \mathbf{y} - \tilde{\mathbf{w}}^\top G^\top \hat{\beta} - \hat{\gamma} \tilde{\mathbf{z}} \quad (\text{C.3a})$$

$$\text{s.t.} \quad O\tilde{\mathbf{w}} + P\tilde{\mathbf{z}} + S\tilde{\mathbf{v}} \preceq_{\mathcal{K}} \mathbf{s} \quad (\text{C.3b})$$

$$\underline{\xi} \leq \tilde{\mathbf{w}} \leq \bar{\xi} \quad (\text{C.3c})$$

$$\tilde{\mathbf{z}} \geq 0. \quad (\text{C.3d})$$

The resulting formulation involves bilinear terms $\tilde{w}_i y_j$ for $i = 1, \dots, I_1, j = 1, \dots, m$ 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 [3] as the bilinear terms can be presented as sums of products between bounded continuous variables $\tilde{w}_i$ and binary variables.

An alternative to CCG is to apply a Benders-type cutting-plane scheme directly to the recourse function viewed as a function of the master variables (β, γ) . Define

$$\Phi(\beta, \gamma) := \max_{\mathbf{y} \in \mathcal{Y}, \tilde{\mathbf{w}}, \tilde{\mathbf{z}}, \tilde{\mathbf{v}}} \left\{ (Q\tilde{\mathbf{w}} + \mathbf{q}_0)^\top \mathbf{y} - (G\tilde{\mathbf{w}})^\top \beta - \tilde{\mathbf{z}}^\top \gamma : O\tilde{\mathbf{w}} + P\tilde{\mathbf{z}} + S\tilde{\mathbf{v}} \preceq_{\mathcal{K}} \mathbf{s}, \underline{\xi} \leq \tilde{\mathbf{w}} \leq \bar{\xi}, \tilde{\mathbf{z}} \geq 0 \right\},$$

which coincides with the subproblem in (C.3). The mapping Φ is convex and piecewise linear in (β, γ) , since it is the pointwise maximum over a finite collection of linear functions indexed by the extreme points $(\hat{\mathbf{y}}, \tilde{\mathbf{w}}^*, \tilde{\mathbf{z}}^*)$ of the subproblem's feasible region. Each such extreme point yields a supporting hyperplane

$$\alpha \geq (Q\tilde{\mathbf{w}}^* + \mathbf{q}_0)^\top \hat{\mathbf{y}} - (G\tilde{\mathbf{w}}^*)^\top \beta - (\tilde{\mathbf{z}}^*)^\top \gamma,$$

which is the Benders cut accumulated by the master.

The Benders master problem is therefore

$$\min_{\alpha, \beta, \gamma} \quad \alpha + \mu^\top \beta + \sigma^\top \gamma \quad (\text{C.4a})$$

$$\text{s.t.} \quad \alpha \geq (Q\tilde{\mathbf{w}}_l^* + \mathbf{q}_0)^\top \hat{\mathbf{y}}_l - (G\tilde{\mathbf{w}}_l^*)^\top \beta - (\tilde{\mathbf{z}}_l^*)^\top \gamma, \quad l \in \mathcal{L} \quad (\text{C.4b})$$

$$\gamma \geq 0, \beta \in \mathbb{R}^I, \alpha \in \mathbb{R}, \quad (\text{C.4c})$$

where $\mathcal{L}$ collects the cuts generated so far and $(\hat{\mathbf{y}}_l, \tilde{\mathbf{w}}_l^*, \tilde{\mathbf{z}}_l^*)$ is the optimal solution returned by the subproblem at iteration l . The Benders master (C.4) carries the variables (α, β, γ) and a single scalar cut per iteration, and each cut is a linear outer approximation.

Both decomposition methods follow the same master–subproblem framework, summarized in Algorithm 1. At each iteration, the master problem produces a candidate dual solution $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ and a lower bound on the optimal value v_0 . The subproblem then searches for a recourse solution that violates the current relaxation. If no violation is found, optimality is certified. Otherwise, the subproblem returns information that is incorporated into the master.

The key difference lies in how this information is used. The CCG algorithm explicitly introduces the newly identified recourse solution and its associated dual variables into the master problem, thereby preserving the structure of the finite convex reformulation. In contrast, the Benders approach adds only a supporting hyperplane to approximate the recourse function. As demonstrated in the computational experiments, retaining the recourse structure through CCG can lead to substantially stronger master problems and faster convergence.

Algorithm 1 Unified master–subproblem scheme (CCG / Benders)

- 1: **Input:** method $\in \{\text{CCG}, \text{BENDERS}\}$, tolerance $\epsilon > 0$, iteration limit $l^{\max}$.
 - 2: **Initialize:** $LB = -\infty$, $\mathcal{L} = \emptyset$, $UB = +\infty$, $l = 0$.
 - 3: **while** $l < l^{\max}$ **do**
 - 4: **if** method = CCG **then**
 - 5: Solve the relaxed master (C.1) restricted to $k \in \mathcal{J}$ to obtain $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$.
 - 6: **else**
 - 7: Solve the Benders master (C.4) to obtain $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$.
 - 8: **end if**
 - 9: Update $LB \leftarrow \hat{\alpha} + \mu^\top \hat{\beta} + \sigma^\top \hat{\gamma}$.
 - 10: Solve the subproblem (C.3) at $(\hat{\beta}, \hat{\gamma})$ to obtain $(\hat{\mathbf{y}}, \tilde{\mathbf{w}}^*, \tilde{\mathbf{z}}^*)$ and value $\hat{v}$; update $UB \leftarrow \min\{UB, \hat{v} + \mu^\top \hat{\beta} + \sigma^\top \hat{\gamma}\}$.
 - 11: **if** $UB - LB \leq \epsilon$ **then**
 - 12: **return** $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$.
 - 13: **else if** method = CCG **then**
 - 14: Add the recourse solution $\hat{\mathbf{y}}$ to the active set $\mathcal{J}$. Introduce new variables $(\delta_{|\mathcal{J}|+1}, \bar{\theta}_{|\mathcal{J}|+1}, \underline{\theta}_{|\mathcal{J}|+1})$, and add to the master the constraints

$$\begin{aligned} \alpha - \mathbf{s}^\top \delta_{|\mathcal{J}|+1} - \bar{\xi}^\top \bar{\theta}_{|\mathcal{J}|+1} - \underline{\xi}^\top \underline{\theta}_{|\mathcal{J}|+1} &\geq \mathbf{q}_0^\top \hat{\mathbf{y}} \\ G^\top \beta + O^\top \delta_{|\mathcal{J}|+1} + \bar{\theta}_{|\mathcal{J}|+1} + \underline{\theta}_{|\mathcal{J}|+1} &= Q^\top \hat{\mathbf{y}} \\ \gamma + P^\top \delta_{|\mathcal{J}|+1} &\geq 0, \quad S^\top \delta_{|\mathcal{J}|+1} = 0 \\ \delta_{|\mathcal{J}|+1} &\geq_{\mathcal{K}} 0, \quad \bar{\theta}_{|\mathcal{J}|+1} \geq 0, \quad \underline{\theta}_{|\mathcal{J}|+1} \leq 0. \end{aligned}$$
 - 15: **else**
 - 16: Append the Benders cut

$$\alpha \geq (Q\tilde{\mathbf{w}}^* + \mathbf{q}_0)^\top \hat{\mathbf{y}} - (G\tilde{\mathbf{w}}^*)^\top \beta - (\tilde{\mathbf{z}}^*)^\top \gamma$$
 to the master, and set $\mathcal{L} \leftarrow \mathcal{L} \cup \{l+1\}$.
 - 17: **end if**
 - 18: $l \leftarrow l + 1$.
 - 19: **end while**
-

Appendix D. 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 $\Omega := \{\hat{\xi}^\omega, \omega = 1, \dots, N\} \subset \mathcal{U}$. Let $\mathbb{P}_0 := 1/N \sum_{\omega=1}^N \delta_{\hat{\xi}^\omega}$ be the empirical distribution, where $\delta_{\hat{\xi}^\omega}$ is the Dirac distribution concentrating unit mass at $\hat{\xi}^\omega$. 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

$$d_W(\mathbb{P}_1, \mathbb{P}_2) := \inf_{\mathbb{Q} \sim (\mathbb{P}_1, \mathbb{P}_2)} \mathbb{E}_{\mathbb{Q}}(\|\xi^1 - \xi^2\|). \quad (\text{D.1})$$

Here, ξ^1, ξ^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 $\xi' = \hat{\xi}^\omega$, $\omega \in \Omega$. The Wasserstein ambiguity set $\mathcal{W}$ is equivalent to

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

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

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

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

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

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

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

Proof. The proof is similar to Wiesemann et al. [8] and Bertsimas et al. [1]. We first show that the marginal distributions of ξ under any $\mathbb{P} \in \mathcal{V}$ belong 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 ξ under $\mathbb{P}$ and $\mathbb{P}_\omega^*$ be the marginal distribution of ξ under $\mathbb{P}_\omega$ for all $\omega \in \Omega$.

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

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

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

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

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

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

$$\text{s.t.} \quad \sum_{k=1}^K u_k^\omega = 1, \quad \omega \in \Omega \quad (\text{D.2b})$$

$$\frac{1}{N} \sum_{\omega \in \Omega} \sum_{k=1}^K z_k^\omega \leq \varepsilon \quad (\text{D.2c})$$

$$O\mathbf{w}_k^\omega + S\mathbf{v}_k^\omega \preceq_{\mathcal{K}} \mathbf{s} u_k^\omega, \quad k \in [K], \quad \omega \in \Omega \quad (\text{D.2d})$$

$$\|\mathbf{w}_k^\omega - u_k^\omega \hat{\boldsymbol{\xi}}^\omega\| \leq z_k^\omega, \quad k \in [K], \quad \omega \in \Omega \quad (\text{D.2e})$$

$$\bar{\boldsymbol{\xi}} u_k^\omega \geq \mathbf{w}_k^\omega \geq \underline{\boldsymbol{\xi}} u_k^\omega, \quad k \in [K], \quad \omega \in \Omega \quad (\text{D.2f})$$

$$u_k^\omega \geq \mathbf{0}, \quad z_k^\omega \geq \mathbf{0}, \quad k \in [K], \quad \omega \in \Omega. \quad (\text{D.2g})$$

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 $K_0 \leq \min\{N+1, K\}$ 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 [2], there exists a worst-case distribution supported on at most $N+1$ points. Denote them as $\boldsymbol{\xi}^k$, $k = 1, \dots, N+1$. 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 $\mathbf{y}$ -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 $K_0 \leq K$. $\square$

Appendix E. CVaR Formulation

In (1), consider the case where the risk measure $\rho_{\mathbb{P}}$ is conditional value-at-risk (CVaR):

$$\sup_{\mathbb{P} \in \mathcal{D}} \text{CVaR}_{\epsilon}^{\mathbb{P}}[f(\boldsymbol{\xi})] = \sup_{\mathbb{P} \in \mathcal{G}} \text{CVaR}_{\epsilon}^{\mathbb{P}}[f(\boldsymbol{\xi})], \quad (\text{E.1})$$

where the equality follows Corollary 1. When $\epsilon = 0$, $\text{CVaR}_{\epsilon}^{\mathbb{P}}[\cdot]$ reduces to the expectation measure $\mathbb{E}_{\mathbb{P}}[\cdot]$.

Problem (E.1) is equivalent to [see, e.g., 4]

$$\min_t \quad t + \frac{1}{1-\epsilon} F(t), \quad (\text{E.2})$$

where $F(t) := \sup_{\mathbb{P} \in \mathcal{G}} \mathbb{E}_{\mathbb{P}}[(f(\boldsymbol{\xi}) - t)^+]$.

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. With a given t , for every realization of $\boldsymbol{\xi} \in \mathcal{U}$ such that $f(\boldsymbol{\xi}) > t$, 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 $k = 1, \dots, K$, let

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

Moreover, for those $\boldsymbol{\xi} \in \mathcal{U}$ such that $f(\boldsymbol{\xi}) \leq t$, we introduce $u_0 = \mathbb{P}(f(\boldsymbol{\xi}) \leq t)$, $\mathbf{w}_0 = \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi} | f(\boldsymbol{\xi}) \leq t] \mathbb{P}(f(\boldsymbol{\xi}) \leq t)$, and $\mathbf{z}_0 = \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi} | f(\boldsymbol{\xi}) \leq t] \mathbb{P}(f(\boldsymbol{\xi}) \leq t)$. Observe that $\sum_{k=0}^K u_k = 1$, $G \sum_{k=0}^K \mathbf{w}_k = \boldsymbol{\mu}$, and $\sum_{k=0}^K \mathbf{z}_k \leq \boldsymbol{\sigma}$.

Due to the convexity of the lifted support set, for positive u_i , we have $(\mathbf{w}_i/u_i, \mathbf{z}_i/u_i) \in \hat{\mathcal{U}}$ which implies $O\mathbf{w}_i + P\mathbf{z}_i + S\mathbf{v}_i \preceq \mathbf{s} u_i$ for some $\mathbf{v}_i$. Following the law of total probability, the expectation of $(f(\boldsymbol{\xi}) - t)^+$ can be expressed as

$$\mathbb{E}_{\mathbb{P}}[(f(\boldsymbol{\xi}) - t)^+] = \sum_{i=k}^K \mathbb{E}_{\mathbb{P}} \left[\mathbf{q}(\boldsymbol{\xi})^\top \mathbf{y}^k - t \mid \hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k \right] \mathbb{P}(\hat{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{y}^k).$$

These observations bring us to the following result.

Theorem 4. Given all the feasible solutions of $\mathcal{Y} = \{\mathbf{y}^k\}_{k=1}^K$, the worst-case expectation $F(t)$ in (E.2) equals to

$$V(t, \mathbf{y}^{[K]}) := \max_{(u_k, \mathbf{w}_k, \mathbf{z}_k, \mathbf{v}_k)_{k=0}^K} \sum_{k=1}^K (Q\mathbf{w}_k + \mathbf{q}_0 u_k)^\top \mathbf{y}^k + t u_0 - t \quad (\text{E.3a})$$

$$s.t. \quad \sum_{k=0}^K u_k = 1 \quad (\text{E.3b})$$

$$G \sum_{k=0}^K \mathbf{w}_k = \boldsymbol{\mu} \quad (\text{E.3c})$$

$$\sum_{k=0}^K \mathbf{z}_k \leq \boldsymbol{\sigma} \quad (\text{E.3d})$$

$$O\mathbf{w}_k + P\mathbf{z}_k + S\mathbf{v}_k \preceq_{\mathcal{K}} s u_k, \quad k \in [K] \cup \{0\} \quad (\text{E.3e})$$

$$\bar{\xi} u_k \geq \mathbf{w}_k \geq \underline{\xi} u_k, \quad k \in [K] \cup \{0\} \quad (\text{E.3f})$$

$$u_k \geq 0, \quad \mathbf{z}_k \geq 0, \quad k \in [K] \cup \{0\}. \quad (\text{E.3g})$$

Constraints (E.3f)-(E.3g) follow the feasibility of the lifted support set $\bar{\mathcal{U}}$. It is clear that the formulation above constitutes a valid convex relaxation of $F(t)$. To show the upper bound is tight, we construct a worst-case distribution that achieves the bound, which is presented in the proof below.

Proof. Given t , we construct a discrete distribution that attains the tight upper bound $V(t, \mathbf{y}^{[K]})$ based on an optimal solution $(\bar{u}_k, \bar{\mathbf{w}}_k, \bar{\mathbf{z}}_k)_{k=1}^K$ of problem (E.3). 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 (ξ, ζ) , 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

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

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

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 the solution $(\bar{u}_k, \bar{\mathbf{w}}_k, \bar{\mathbf{z}}_k)_{k=1}^K$

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

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 $f(\bar{\mathbf{w}}_k/\bar{u}_k)$, and $(f(\bar{\mathbf{w}}_k/\bar{u}_k) - t)^+ \geq f(\bar{\mathbf{w}}_k/\bar{u}_k) - t$, for all $k \in [K] \cup \{0\}$ such that $\bar{u}_k > 0$. The second inequality follows from the observation $\bar{\mathbb{P}} \in \mathcal{G}$. $\square$

Appendix F. Proof of Theorem 3

Proof. It is clear that the formulation (4) constitutes a valid relaxation of $\sup_{\mathbb{P} \in \mathcal{G}} \mathbb{P}(f(\xi) \geq \tau)$. 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 (ξ, ζ) , 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

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

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

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

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

Appendix G. Multi-item Newsvendor Problem: More Results

To assess the practical scalability of the proposed decomposition algorithms, we benchmark the CCG algorithm and Benders-type cutting-plane method against the full reformulation (3) on the multi-item newsvendor instance introduced in Section 4. We consider instances with the number of items $N \in \{2, 4, 8, 16, 32\}$. For each N , we generate 10 instances by varying the item prices and report the average runtime or the optimality gap. All experiments are conducted with a time limit of 1000 seconds and a convergence tolerance of 10^{-3} . Results are summarized in Table G.4.

Table G.4: Computational comparison of CCG, Benders, and the full formulation on the newsvendor DRO instances. (Time limit: 1000s)

#Items	CCG		Benders		Full Formulation	
	gap	time (s)	gap	time (s)	gap	time (s)
2	–	0.08	–	0.46	–	0.01
4	–	0.38	–	3.43	–	0.14
8	–	8.57	–	149.23	Fail [‡]	
16	–	80.26	0.95%	1000.00	Fail [‡]	
32	–	995.43	58.49%	1000.02	Fail [‡]	

[‡] Gurobi failed at presolve; no gap returned.

The full reformulation (C.1) is the fastest approach for the smallest instances ($N \leq 4$), where explicit enumeration of recourse solutions remains manageable. However, its size grows rapidly with the number of items because the number of feasible recourse solutions satisfies $(|\mathcal{Y}| = 6^N)$ and thus it fails during presolve when $N \geq 8$.

Both decomposition methods avoid this combinatorial growth by generating recourse solutions when needed. Among them, the CCG method consistently outperforms the Benders approach. For example, when $N = 16$ or 32, the Benders method fails to converge within the time limit, whereas CCG solves all of them. This result is consistent with the theoretical results developed in Section 2.

Although the newsvendor instances considered here are not identical to the setting used to develop the decomposition methods in Appendix C, where the first-stage decision $\mathbf{x}$ is treated as fixed, the same framework applies without modification. Here, the first-stage order quantity $\mathbf{x}$ is a decision variable; nevertheless, the proposed decomposition algorithms naturally extend to this setting and remains computationally effective. To gain further insights into its performance, we track the critical recourse solutions throughout the decomposition process.

At iteration l , after solving the subproblem (C.3), a recourse solution is added to the accumulated pool $\mathcal{J}^{(l)}$ if the current first-stage solution is not optimal. We then solve the primal reformulation (3) with $\mathcal{J}^{(l)}$ and record the number of critical recourse solutions with nonzero probability in the current worst-case distribution, which can be viewed as an estimate of the support size K_0 appearing in Theorem 2. Figure G.4 reports these quantities for an 8-item instance.

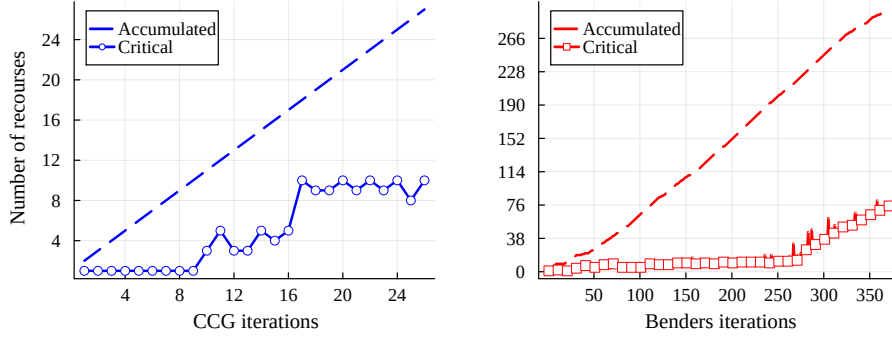


Figure G.4: Accumulated vs. critical recourses over CCG (left) and Benders (right) iterations for $N = 8$. Accumulated: all distinct recourse solutions $\mathbf{y}^k$ collected so far; Critical: number of $\mathbf{y}^k$ with $u_k > 0$ under the worst-case distribution on the current accumulated ones.

We observe that CCG converges in substantially fewer iterations (27) than Benders (371). At convergence, CCG has also collected fewer distinct $\mathbf{y}$ -solutions (27) than Benders (75) did. This superior efficiency of the proposed CCG algorithm can be attributed to two reasons. First, CCG identifies the $\mathbf{y}$ -solutions that are most critical under the worst-case distribution, adding a new solution at every iteration; in contrast, Benders repeatedly regenerates $\mathbf{y}$ -solutions that have already been collected. Second, CCG utilizes each identified $\mathbf{y}$ -solution more informatively by augmenting the master with multiple variables and constraints per iteration, whereas Benders appends only a single cut.

Appendix H. Distributionally Robust Chance-Constrained Network Interdiction

This section studies the two-stage distributionally robust chance-constrained program with integer recourse, where the chance constraint bounds the worst-case probability in the form of $\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{P}(f_{\mathbf{x}}(\xi) \geq \tau) \leq \epsilon$. Following the finite reformulation results for probability-based risk measures developed in Section 3, we derive a conic reformulation and a CCG algorithm for efficiently solving it. We then apply the proposed reformulation and algorithm to a shortest-path network interdiction problem to demonstrate their effectiveness and efficiency.

Appendix H.1. Generic Formulation and Decomposition

Consider the following two-stage distributionally robust chance-constrained program (TSDR-CC) with integer recourse

$$\min_{\mathbf{x} \in \mathcal{X}} \left\{ \mathbf{c}^\top \mathbf{x} : \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{P}(f_{\mathbf{x}}(\xi) \geq \tau) \leq \epsilon \right\}, \quad (\text{H.1})$$

where $\mathbf{x}$ is a first-stage decision, $\epsilon \in (0, 1)$ is a risk tolerance, and τ is a predetermined target. The recourse function is given by the binary program $f_{\mathbf{x}}(\xi) = \max_{\mathbf{y} \in \mathcal{Y}} \mathbf{q}_{\mathbf{x}}(\xi)^\top \mathbf{y}$ with $\mathcal{Y} \subset \{0, 1\}^m$.

The objective coefficient depends on the first-stage decision through $\mathbf{q}_{\mathbf{x}}(\xi) = \mathbf{h}(\xi) + T(\xi)\mathbf{x}$, where $\mathbf{h}(\xi) = H\xi + \mathbf{h}_0$ and $T(\xi) = T_0 + \sum_{n=1}^{I_1} \xi_n T_n$ are both linear in ξ . Writing $T(\mathbf{x}, \mathbf{y}) := [\mathbf{x}^\top T_n^\top \mathbf{y}]_{n=1}^{I_1}$, the recourse objective for a fixed $\mathbf{y}$ is $\mathbf{q}_{\mathbf{x}}(\xi)^\top \mathbf{y} = \xi^\top (H^\top \mathbf{y} + T(\mathbf{x}, \mathbf{y})) + (\mathbf{h}_0 + T_0 \mathbf{x})^\top \mathbf{y}$.

Let $\mathcal{Y} = \{\mathbf{y}^k\}_{k=1}^K$. By conditioning on the optimal recourse solution, Theorem 3 implies that the worst-case probability

$$U(\mathbf{x}) = \max_{(u_k, \mathbf{w}_k, \mathbf{z}_k, \mathbf{v}_k)_{k=0}^K} \sum_{k=1}^K u_k \quad (\text{H.2a})$$

$$\text{s.t.} \quad \mathbf{w}_k^\top (H^\top \mathbf{y}^k + T(\mathbf{x}, \mathbf{y}^k)) + u_k(\mathbf{h}_0 + T_0 \mathbf{x})^\top \mathbf{y}^k \geq \tau u_k, \quad k \in [K] \quad (\text{H.2b})$$

$$\sum_{k=0}^K u_k = 1 \quad (\text{H.2c})$$

$$G \sum_{k=0}^K \mathbf{w}_k = \boldsymbol{\mu} \quad (\text{H.2d})$$

$$\sum_{k=0}^K \mathbf{z}_k \leq \boldsymbol{\sigma} \quad (\text{H.2e})$$

$$O\mathbf{w}_k + P\mathbf{z}_k + S\mathbf{v}_k \preceq_{\mathcal{K}} \mathbf{s} u_k, \quad k \in [K] \cup \{0\} \quad (\text{H.2f})$$

$$\underline{\xi} u_k \leq \mathbf{w}_k \leq \bar{\xi} u_k, \quad k \in [K] \cup \{0\} \quad (\text{H.2g})$$

$$u_k \geq 0, \quad \mathbf{z}_k \geq 0, \quad k \in [K] \cup \{0\}. \quad (\text{H.2h})$$

Introducing dual variables $\pi_k \geq 0$, $\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma} \geq 0$, $\boldsymbol{\delta}_k \succeq_{\mathcal{K}} 0$, $\bar{\boldsymbol{\theta}}_k \geq 0$, and $\underline{\boldsymbol{\theta}}_k \leq 0$ associated with constraints (H.2b)–(H.2g), the conic dual of (H.2) is

$$U(\mathbf{x}) = \min_{\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \pi, \boldsymbol{\delta}, \bar{\boldsymbol{\theta}}, \underline{\boldsymbol{\theta}}} \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma} \quad (\text{H.3a})$$

$$\text{s.t.} \quad \alpha \geq 1 + \pi_k((\mathbf{h}_0 + T_0 \mathbf{x})^\top \mathbf{y}^k - \tau) + \mathbf{s}^\top \boldsymbol{\delta}_k + \bar{\boldsymbol{\xi}}^\top \bar{\boldsymbol{\theta}}_k + \underline{\boldsymbol{\xi}}^\top \underline{\boldsymbol{\theta}}_k, \quad k \in [K] \quad (\text{H.3b})$$

$$\alpha \geq \mathbf{s}^\top \boldsymbol{\delta}_0 + \bar{\boldsymbol{\xi}}^\top \bar{\boldsymbol{\theta}}_0 + \underline{\boldsymbol{\xi}}^\top \underline{\boldsymbol{\theta}}_0 \quad (\text{H.3c})$$

$$G^\top \boldsymbol{\beta} + O^\top \boldsymbol{\delta}_k + \bar{\boldsymbol{\theta}}_k + \underline{\boldsymbol{\theta}}_k = \pi_k (H^\top \mathbf{y}^k + T(\mathbf{x}, \mathbf{y}^k)), \quad k \in [K] \quad (\text{H.3d})$$

$$G^\top \boldsymbol{\beta} + O^\top \boldsymbol{\delta}_0 + \bar{\boldsymbol{\theta}}_0 + \underline{\boldsymbol{\theta}}_0 = 0 \quad (\text{H.3e})$$

$$\boldsymbol{\gamma} + P^\top \boldsymbol{\delta}_k \geq 0, \quad k \in [K] \cup \{0\} \quad (\text{H.3f})$$

$$S^\top \boldsymbol{\delta}_k = 0, \quad k \in [K] \cup \{0\} \quad (\text{H.3g})$$

$$\pi_k \geq 0, \quad \boldsymbol{\gamma} \geq 0, \quad \boldsymbol{\delta}_k \succeq_{\mathcal{K}} 0, \quad \bar{\boldsymbol{\theta}}_k \geq 0, \quad \underline{\boldsymbol{\theta}}_k \leq 0, \quad k \in [K] \cup \{0\}. \quad (\text{H.3h})$$

Under the Slater's condition in Assumption 2, strong duality holds between the primal (H.2) and the dual (H.3). Replacing the chance constraint $U(\mathbf{x}) \leq \epsilon$ with the dual representation (H.3), the chance-constrained problem (H.1) is equivalent to the mixed-integer conic program

$$\min_{\mathbf{x} \in \mathcal{X}, (\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \pi, \boldsymbol{\delta}, \bar{\boldsymbol{\theta}}, \underline{\boldsymbol{\theta}})} \left\{ \mathbf{c}^\top \mathbf{x} : \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta} + \boldsymbol{\sigma}^\top \boldsymbol{\gamma} \leq \epsilon, (\text{H.3b})\text{--}(\text{H.3g}) \right\}, \quad (\text{H.4})$$

which we refer to as the full formulation.

When only a subset $\mathcal{J} \subseteq [K]$ of $\mathbf{y}$ -solutions is available, imposing constraints (H.3b)–(H.3g) with $k \in \mathcal{J}$ yields a relaxed master problem whose optimal value lower-bounds that of the full formulation (H.4). This observation motivates a CCG approach, following a similar master–subproblem scheme as in the expectation case of Appendix C. Given an optimal solution $(\hat{\mathbf{x}}, \hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ to the relaxed master problem, we seek a $\mathbf{y}$ -solution not included in the set $\mathcal{J}$ whose dual constraint (H.3b) is violated. For a fixed solution $(\hat{\mathbf{x}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$, minimizing the right-hand side of (H.3b) over

its associated block of dual variables $(\pi, \delta, \bar{\theta}, \underline{\theta})$ and maximizing over $\mathbf{y}$ yields the separation problem

$$S(\hat{\mathbf{x}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \max_{\mathbf{y} \in \mathcal{Y}} \min_{\pi, \delta, \bar{\theta}, \underline{\theta}} \quad 1 + \pi((\mathbf{h}_0 + T_0 \hat{\mathbf{x}})^\top \mathbf{y} - \tau) + \mathbf{s}^\top \delta + \bar{\boldsymbol{\xi}}^\top \bar{\boldsymbol{\theta}} + \underline{\boldsymbol{\xi}}^\top \underline{\boldsymbol{\theta}} \quad (\text{H.5a})$$

$$\text{s.t.} \quad G^\top \hat{\boldsymbol{\beta}} + O^\top \delta + \bar{\boldsymbol{\theta}} + \underline{\boldsymbol{\theta}} = \pi(H^\top \mathbf{y} + T(\hat{\mathbf{x}}, \mathbf{y})) \quad (\text{H.5b})$$

$$\hat{\boldsymbol{\gamma}} + P^\top \delta \geq 0 \quad (\text{H.5c})$$

$$S^\top \delta = 0 \quad (\text{H.5d})$$

$$\pi \geq 0, \delta \succeq_{\mathcal{K}} 0, \bar{\boldsymbol{\theta}} \geq 0, \underline{\boldsymbol{\theta}} \leq 0. \quad (\text{H.5e})$$

Introducing conic dual variables $\tilde{\mathbf{w}}, \tilde{\mathbf{z}}, \tilde{\mathbf{v}}$ for the constraints of the inner minimization and taking its dual, following the same procedure as in the CVaR case, the separation problem (H.5) is equivalent to

$$S(\hat{\mathbf{x}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \max_{\mathbf{y} \in \mathcal{Y}, \tilde{\mathbf{w}}, \tilde{\mathbf{z}}, \tilde{\mathbf{v}}} \quad 1 - (G\tilde{\mathbf{w}})^\top \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\gamma}}^\top \tilde{\mathbf{z}} \quad (\text{H.6a})$$

$$\text{s.t.} \quad \tilde{\mathbf{w}}^\top (H^\top \mathbf{y} + T(\hat{\mathbf{x}}, \mathbf{y})) + (\mathbf{h}_0 + T_0 \hat{\mathbf{x}})^\top \mathbf{y} \geq \tau \quad (\text{H.6b})$$

$$O\tilde{\mathbf{w}} + P\tilde{\mathbf{z}} + S\tilde{\mathbf{v}} \preceq_{\mathcal{K}} \mathbf{s} \quad (\text{H.6c})$$

$$\underline{\boldsymbol{\xi}} \leq \tilde{\mathbf{w}} \leq \bar{\boldsymbol{\xi}}, \tilde{\mathbf{z}} \geq 0. \quad (\text{H.6d})$$

Let $\hat{S}$ denote the optimal value of problem (H.6) and $\hat{\mathbf{y}}$ an optimal solution. If $\hat{\alpha} \geq \hat{S}$, then $(\hat{\mathbf{x}}, \hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ is optimal to the chance-constrained problem (H.1). Otherwise, the recourse solution $\hat{\mathbf{y}}$ is added to the set $\mathcal{J}$ and the relaxed master problem is resolved. The overall procedure is summarized in Algorithm 2.

Algorithm 2 Column-and-constraint generation for the chance-constrained problem (H.1)

- 1: **Initialize:** a subset $\mathcal{J} \subseteq [K]$ with a single recourse solution; tolerance $\epsilon > 0$.
 - 2: **while** true **do**
 - 3: Solve the relaxed master (H.4) restricted to $k \in \mathcal{J}$ to obtain $(\hat{\mathbf{x}}, \hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ and the lower bound $\mathbf{c}^\top \hat{\mathbf{x}}$.
 - 4: Solve the subproblem (H.6) at $(\hat{\mathbf{x}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ to obtain its value $\hat{S}$ and a maximizer $\hat{\mathbf{y}}$.
 - 5: **if** $\hat{S} \leq \hat{\alpha} + \epsilon$ **then return** $\hat{\mathbf{x}}$; **else** set $\mathcal{J} \leftarrow \mathcal{J} \cup \{\hat{\mathbf{y}}\}$.
 - 6: **end while**
-

Appendix H.2. Network Interdiction under DRO Chance Constraints

We instantiate (H.1) as a risk-averse shortest-path interdiction problem [7]. Consider a directed graph $G = (V, E)$ with source node s and sink node t . The interdictor selects a set of arcs to interdict, represented by binary variables $\mathbf{x} \in \{0, 1\}^{|E|}$, incurring a first-stage cost c_e for each interdicted arc e . After the uncertain arc costs $\boldsymbol{\xi}$ are realized, a network user chooses the shortest s - t path. Traversing an interdicted arc e incurs an additional penalty d_e .

Let $\mathcal{Y}$ denote the set of feasible s - t paths, the recourse problem is written as

$$f_{\mathbf{x}}(\boldsymbol{\xi}) = \max_{\mathbf{y} \in \mathcal{Y}} \sum_e (-\xi_e - d_e x_e) y_e, \quad (\text{H.7})$$

where $f_{\mathbf{x}}(\boldsymbol{\xi})$ represents the negative of the shortest-path cost under the interdiction decision $\mathbf{x}$. The arc base costs ξ_e are uncertain and modeled using a marginal-moment ambiguity set, while the interdiction penalties d_e are deterministic. The interdictor seeks a solution such that the shortest-path cost $-f_{\mathbf{x}}(\boldsymbol{\xi})$ exceeds a prescribed threshold $-\tau$ with high probability under all distributions in the ambiguity set $\mathcal{D}$. Equivalently, the probability that $f_{\mathbf{x}}(\boldsymbol{\xi})$ exceeds τ remains below the risk tolerance ϵ in the worst case over $\mathcal{D}$. The TSDR-CC problem is given as

$$\min_{\mathbf{x}} \left\{ \mathbf{c}^\top \mathbf{x} : \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{P}(f_{\mathbf{x}}(\boldsymbol{\xi}) \geq \tau) \leq \epsilon \right\}. \quad (\text{H.8})$$

Test instances. Following the grid network setting of Song and Shen [7], consider an $n \times n$ grid with node set V ($|V| = n^2$) and bidirectional arcs between adjacent nodes, resulting in $|E| = 4n(n - 1)$ arcs. The source node s and sink node t are located at opposite corners of the grid. For each arc, the nominal cost is sampled from $\bar{\xi}_e \sim U(1, 10)$, the interdiction penalty from $d_e \sim U(1, 5)$, and the interdiction cost from $c_e \sim U(1, 5)$. The ambiguity set is constructed using the marginal moments and support information: $\mu_e = \bar{\xi}_e$, $\sigma_e^2 = (\bar{\xi}_e/3)^2$, $\xi_e \in [\bar{\xi}_e/3, 5\bar{\xi}_e/3]$. The target threshold is set to $\tau = -0.5L$, where L is the shortest-path length under the mean arc costs without interdiction.

Appendix H.3. Experiment 1: Computational Performance

We solve the problem with grid networks of size $n \in \{5, 6, 7\}$ and risk levels $\epsilon \in \{0.1, 0.2\}$. For each setting, we generate three random instances and solve them with a 1000-second time limit and an optimality threshold of 10^{-4} . Table H.5 reports the average solution time (AvT), average number of CCG iterations (AvIters), and average optimal interdiction cost (AvCost), computed over the instances solved to optimality. The column "#opt" reports the number of instances solved within the time limit.

Table H.5: Computational performance of the CCG method on grid network instances ($\tau = -0.5L$).

Instance	ϵ	#opt	AvT (s)	AvIters	AvCost
5×5	0.1	3/3	8.0	6.7	12.3
	0.2	3/3	29.8	9.0	14.0
6×6	0.1	3/3	35.0	8.3	11.0
	0.2	3/3	99.0	9.0	11.0
7×7	0.1	2/3	268.2	17.5	11.5
	0.2	2/3	466.4	14.0	11.5

The proposed CCG algorithm solves all 5×5 and 6×6 instances within the time limit. For the larger 7×7 networks, two out of three instances are solved to optimality for each risk level. As expected, the solution time increases with network size due to the growing complexity of the interdiction and recourse decisions. The runtime also tends to increase as the risk tolerance ϵ becomes less restrictive. Nevertheless, the algorithm converges in a relatively small number of iterations across all tested instances, suggesting that only a limited number of critical recourse solutions are needed to certify optimality.

Appendix H.4. Experiment 2: Cost–Risk Sensitivity

Fixing a single 5×5 instance, we examine the relationship between interdiction cost and risk tolerance by varying the risk level $\epsilon \in \{0.05, 0.10, \dots, 0.30\}$ for three target thresholds $\tau \in \{-0.4L, -0.45L, -0.5L\}$. Figure H.5 reports the resulting optimal interdiction costs.

As expected, the optimal interdiction cost is nonincreasing in ϵ . Allowing a larger worst-case probability of violating the target threshold relaxes the chance constraint and reduces the amount of interdiction. For a fixed risk level, larger values of τ , lead to higher interdiction costs, since they impose more stringent requirements on the shortest-path length. We also observe a threshold effect: for each value of τ , the optimal cost remains unchanged for small values of ϵ , indicating that the same interdiction strategy remains optimal. Once ϵ exceeds 0.2, the chance constraint becomes sufficiently relaxed to permit less costly interdiction plans, resulting in a noticeable reduction in the optimal cost.

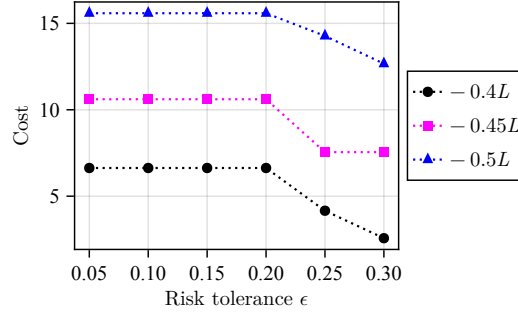


Figure H.5: Optimal interdiction cost versus risk tolerance ϵ on a 5×5 grid instance. Each curve corresponds to a target threshold τ .

References

- [1] Bertsimas, D., Sim, M., Zhang, M., 2019. Adaptive distributionally robust optimization. *Management Science* 65, 604–618.
- [2] Gao, R., Kleywegt, A., 2023. Distributionally robust stochastic optimization with wasserstein distance. *Mathematics of Operations Research* 48, 603–655.
- [3] McCormick, G.P., 1976. Computability of global solutions to factorable nonconvex programs: Part I-Convex underestimating problems. *Mathematical programming* 10, 147–175.
- [4] Rockafellar, R.T., Uryasev, S., et al., 2000. Optimization of conditional value-at-risk. *Journal of risk* 2, 21–42.
- [5] Rogosinski, W.W., 1958. Moments of non-negative mass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 245, 1–27.
- [6] Shapiro, A., 2001. On duality theory of conic linear problems, in: *Semi-Infinite Programming*. Springer, pp. 135–165.
- [7] Song, Y., Shen, S., 2016. Risk-averse shortest path interdiction. *INFORMS Journal on Computing* 28, 527–539.
- [8] Wiesemann, W., Kuhn, D., Sim, M., 2014. Distributionally robust convex optimization. *Operations Research* 62, 1358–1376.