Distributionally Robust Optimization with Decision-Dependent Polyhedral Ambiguity

Hamed Rahimian*[1] and Sanjay Mehrotra†[2]

[1]Department of Industrial Engineering, Clemson University, Clemson SC 29634, USA
[2]Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston IL 60208, USA

Abstract

We consider a two-stage stochastic program with continuous recourse, where the distribution of the random parameters depends on the decisions. Assuming a finite sample space, we study a distributionally robust approach to this problem, where the decision-dependent distributional ambiguity is modeled with a polyhedral ambiguity set. We consider cases where the recourse function and the ambiguity set are either generic or have a special convex/nonconvex structure. We reformulate the resulting problem as a nonconvex two-stage stochastic program, including a bilinearly-constrained bilinear program and a concave minimization problem. We propose finitely-convergent decomposition-based cutting plane algorithms to solve the resulting problems optimally (or near-optimally). The proposed algorithm may also be used to solve two-stage stochastic programs with a random decision-dependent recourse matrix (i.e., bilinear stochasticity on the left-hand side) or a bilinear objective function. We illustrate computational comparative results for joint pricing and stocking decisions on a stylized multiproduct newsvendor problem with price-dependent demand.

Keywords: Distributionally robust optimization Decision-dependent uncertainty Nonconvex optimization Bilinear stochasticity

1 Introduction

Many problems arising in various domains, such as transportation [49] and defense [59, 61], require decisions made before uncertain parameters are realized. Stochastic programming (SP) [14, 65] and robust optimization (RO) [8, 11] are two common frameworks to handle such problems, assuming a full distributional information and only support set information, respectively. Given that a decision-maker may have some partial distributional information about uncertain parameters, a modeling paradigm, referred to as *distributionally robust optimization* (DRO), has recently attracted much attention [6, 33, 54]. DRO unifies SP and RO and protects the decision-maker from ambiguity in the underlying probability. Despite theoretical and algorithmic advances in SP, RO, and DRO, most research often assumes (i) uncertain parameters (or their distributions) are known a priori or belong to a set and (ii) are *exogenous*. In other words, uncertain parameters or their distributions are *decision-independent*. However, these assumptions rarely hold in practice, and uncertain parameters may endogenously depend on the decisions.

*hrahimi@clemson.edu
†mehrotra@northwestern.edu

In this paper, we study a DRO framework for decision-making under uncertainty, where decisions impact the probability distribution of uncertain parameters. More specifically, we consider a DRO problem with a *decision-dependent* ambiguity set as follows:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{P\in\mathcal{P}(\boldsymbol{x})} \mathbb{E}_P\left[h(\boldsymbol{x},\boldsymbol{\xi})\right], \tag{1}$$

where $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ is the decision vector in a nonempty and bounded mixed-integer deterministic feasible region $\mathcal{X} := \{\boldsymbol{x} \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n-n_1} \,|\, \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{d}, \boldsymbol{x} \geq \boldsymbol{0}\}$. We define a random vector $\boldsymbol{\xi} \in \mathbb{R}^d$ on a measurable space $(\Xi, \mathcal{F})$, where $\Xi \subset \mathbb{R}^d$ is the support of $\boldsymbol{\xi}$ and $\mathcal{F}$ is a Borel σ-algebra. Moreover, $h(\boldsymbol{x},\boldsymbol{\xi}) : \mathcal{X} \times \Xi \mapsto \mathbb{R}$ is a random cost function. For a given $\boldsymbol{x} \in \mathcal{X}$, $\mathcal{P}(\boldsymbol{x})$ is a decision-dependent ambiguity set of probability distributions, with P as an element of this set. We have $\mathcal{P}(\boldsymbol{x}) \subseteq \mathfrak{P}(\Xi, \mathcal{F})$, where $\mathfrak{P}(\Xi, \mathcal{F})$ is the set of all probability distributions defined on $(\Xi, \mathcal{F})$. For a fixed $\boldsymbol{x} \in \mathcal{X}$, the inner maximization problem in (1) finds the worst-case expected cost over $\mathcal{P}(\boldsymbol{x})$.

Problem (1) contains a wide range of problems, including decision-dependent distributionally robust two-stage mixed-integer stochastic programs with continuous recourse, where

$$h(\boldsymbol{x},\boldsymbol{\xi}) = \min_{\boldsymbol{y}\in\mathcal{Y}(\boldsymbol{x},\boldsymbol{\xi})} \psi_0(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\xi}), \tag{2}$$

and

$$\mathcal{Y}(\boldsymbol{x},\boldsymbol{\xi}) := \{\boldsymbol{y} \in \mathbb{R}^m \,|\, \boldsymbol{D}\boldsymbol{y} \geq \boldsymbol{B}\boldsymbol{x} + \boldsymbol{b},\ \boldsymbol{y} \geq \boldsymbol{0}\}, \tag{3}$$

with matrices $\boldsymbol{D}$ and $\boldsymbol{B}$, and vector $\boldsymbol{b}$ of appropriate dimensions. Problem (2) involves uncertain parameters in both the objective function and feasible region (3). In this paper, we are especially interested in problems where

$$\psi_0(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\xi}) = \boldsymbol{c}^\top\boldsymbol{x} + \boldsymbol{q}^\top\boldsymbol{y}, \tag{4}$$

or

$$\psi_0(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\xi}) = \boldsymbol{c}^\top\boldsymbol{x} + \boldsymbol{q}^\top\boldsymbol{y} + \boldsymbol{x}^\top\boldsymbol{L}\boldsymbol{y}. \tag{5}$$

Here, $\boldsymbol{\xi}$ includes vectors $\boldsymbol{c}$, $\boldsymbol{q}$, $\boldsymbol{b}$ and matrices $\boldsymbol{D}$, $\boldsymbol{B}$, and $\boldsymbol{L}$. Note that for a fixed $\boldsymbol{\xi} \in \Xi$, (4) and (5) result in a convex and nonconvex recourse function, respectively. We assume that

A1 (Finite Sample Space) Each distribution $P \in \mathcal{P}(\boldsymbol{x})$ has a decision-independent finite support $\Xi = \{\boldsymbol{\xi}_\omega\}_{\omega=1}^N$, for all $\boldsymbol{x} \in \mathcal{X}$, where N is the fixed number of scenarios.

A2 (Complete Recourse) For $\boldsymbol{x} \in \mathbb{R}^n$ and $\omega \in [N]$, we have a complete recourse for problem $h(\boldsymbol{x},\boldsymbol{\xi}_\omega)$, defined in (2), and $\mathcal{Y}(\boldsymbol{x},\boldsymbol{\xi}_\omega)$, defined in (3), is a bounded set.

Remark 1. *Assumption* **(A2)** *implies relatively complete recourse; that is, $\mathcal{Y}(\boldsymbol{x},\boldsymbol{\xi}_\omega)$ is feasible for any $\boldsymbol{x} \in \mathcal{X}$. Thus, $h(\boldsymbol{x},\boldsymbol{\xi}_\omega)$ is bounded from above and $\left\{\boldsymbol{\pi} \geq \boldsymbol{0} \,\middle|\, \boldsymbol{\pi}^\top\boldsymbol{D}_\omega \leq \boldsymbol{0}^\top,\ \boldsymbol{\pi}^\top(\boldsymbol{B}_\omega\boldsymbol{x} + \boldsymbol{b}_\omega) \geq \boldsymbol{0}\right\} = \{\boldsymbol{0}\}$ by Farkas Lamma. Hence, an optimal solution to the corresponding dual problem of (2) is attained at an extreme point; showing that $h(\boldsymbol{x},\boldsymbol{\xi}_\omega)$ is bounded from below (by weak duality). In addition, given that we have a complete recourse for every $\boldsymbol{x} \in \mathbb{R}^n$, we must have $\left\{\boldsymbol{\pi} \geq \boldsymbol{0} \,\middle|\, \boldsymbol{\pi}^\top\boldsymbol{D}_\omega \leq \boldsymbol{0}^\top\right\} = \{\boldsymbol{0}\}$; hence, the set of optimal dual solutions is bounded.*

Given Assumption **(A1)**, we have $\mathcal{F} = 2^\Xi$. For any $P \in \mathfrak{P}(\Xi, \mathcal{F})$, we let $\{p_1, \ldots, p_N\}$ be the probabilities of the corresponding elementary events. As a shorthand notation, we might use $\boldsymbol{p} = [p_1, \ldots, p_N]^T \in \mathbb{R}^N$.

1.1 Decision-Dependent Ambiguity Set of Probability Distributions

In this paper, we consider a general polyhedral decision-dependent ambiguity set of probability distributions, although there are various ways to model the distributional ambiguity. Following the terminology in [54], there are four main groups of ambiguity sets studied in the literature: (i) discrepancy-based ambiguity sets, (ii) moment-based ambiguity sets, (iii) shape-preserving ambiguity sets, and (iv) kernel-based ambiguity sets. A discrepancy-based ambiguity set is constructed by considering all probability distributions that are close to a nominal probability distribution in the sense of a discrepancy measure, e.g., probability metric. Optimal transport discrepancy, see, e.g., [15, 22, 23, 24, 30, 35, 40, 75], and ϕ-divergences [6, 7] with its special cases such as total variation distance, see, e.g., [31, 50, 51, 52, 63], are well studied. Moment-based models contain all probability distributions whose moments satisfy certain properties, especially the first two moments, see, e.g., [19, 62, 73]. For a more in-depth review of DRO and various ambiguity sets, interested readers are referred to [54].

Instead of focusing on a particular type of ambiguity set, e.g., a discrepancy-based or moment-based model, we consider a polyhedral ambiguity set, which subsumes some classes of both models. More precisely, we study a decision-dependent ambiguity set with a *generalized moment and measure inequalities* [54] as follows:

$$\mathcal{P}(\boldsymbol{x}) := \left\{ \boldsymbol{p} \geq \boldsymbol{0} \,\middle|\, \begin{array}{l} \underline{p}_\omega \leq p_\omega \leq \overline{p}_\omega, \ \omega \in [N], \\ \sum_{\omega \in [N]} p_\omega g^i(\boldsymbol{\xi}_\omega) \leq \vartheta^i(\boldsymbol{x}), \ i \in [s] \end{array} \right\}, \tag{6}$$

for $\boldsymbol{x} \in \mathcal{X}$, where $g^i : \Xi \mapsto \mathbb{R}$, $i \in [s]$, with $s \geq 1$. We might use $\boldsymbol{g} = [g^1, \ldots, g^s]$ and $\boldsymbol{\vartheta}(\boldsymbol{x}) = [\vartheta^1(\boldsymbol{x}), \ldots, \vartheta^s(\boldsymbol{x})]$ as a shorthand notation. The first set of constraints in (6) enforce a preference relationship between probability distributions, i.e., lower and upper bounds on the probabilities. To ensure that $\boldsymbol{p}$ is a probability distribution, we set $\vartheta^1(\boldsymbol{x}) = 1$, $g^1 = 1$, $\vartheta^2(\boldsymbol{x}) = -1$, and $g^2 = -1$, in the above definition of $\mathcal{P}(\boldsymbol{x})$, for all $\boldsymbol{x} \in \mathcal{X}$. The authors in [64] proposed the decision-independent version of (6), and its special cases appear in [3, 12, 28, 38, 45, 46], among others. We further assume that

A3 **(Nonempty Ambiguity Set)** For a fixed $\boldsymbol{x} \in \mathcal{X}$, the ambiguity set $\mathcal{P}(\boldsymbol{x})$, defined in (6), is nonempty.

Examples of Ambiguity Sets

A polyhedral ambiguity set in the form of (6) subsumes a wide range of ambiguity sets studied in the DRO literature. An ambiguity set based on Wasserstein distance of order 1, or Kantorovich distance, is represented as:

$$\mathcal{P}(\boldsymbol{x}) = \left\{ \boldsymbol{p} \geq \boldsymbol{0} \,\middle|\, \begin{array}{l} \exists \boldsymbol{z} \text{ s.t. } \sum\limits_{\omega \in [N]} \sum\limits_{\omega' \in [N]} z_{\omega\omega'} \|\boldsymbol{\xi}_\omega - \boldsymbol{\xi}_{\omega'}\| \leq \tau(\boldsymbol{x}), \ \sum\limits_{\omega \in [N]} p_\omega = 1, \\ \sum\limits_{\omega \in [N]} z_{\omega\omega'} = p_\omega, \ \sum\limits_{\omega' \in [N]} z_{\omega\omega'} = q_\omega, \ \omega \in [N] \end{array} \right\},$$

where $\tau(\boldsymbol{x})$ denotes a decision-dependent radius on the Wasserstein distance between a candidate probability distribution $\boldsymbol{p}$ and an empirical probability distribution $\boldsymbol{q}$.

A total variation ambiguity set is represented as:

$$\mathcal{P}(\boldsymbol{x}) = \left\{ \boldsymbol{p} \geq \boldsymbol{0} \,\middle|\, \tfrac{1}{2} \sum_{\omega \in [N]} |p_\omega - q_\omega| \leq \tau(\boldsymbol{x}), \ \sum_{\omega \in [N]} p_\omega = 1 \right\},$$

where $\tau(\boldsymbol{x})$ denotes a decision-dependent radius on the total variation distance between a candidate probability distribution $\boldsymbol{p}$ and an empirical probability distribution $\boldsymbol{q}$.

A moment-based ambiguity set may be represented as:

$$\mathcal{P}(\boldsymbol{x}) = \left\{ \boldsymbol{p} \geq \boldsymbol{0} \;\middle|\; \begin{array}{l} (1-\tau_1)\boldsymbol{\mu}_0(\boldsymbol{x}) \leq \sum\limits_{\omega\in[N]} p_\omega \boldsymbol{\xi}_\omega \leq (1+\tau_1)\boldsymbol{\mu}_0(\boldsymbol{x}), \\ \underline{\tau}_2\big(\boldsymbol{\sigma}_0(\boldsymbol{x})^2 + \boldsymbol{\mu}_0(\boldsymbol{x})^2\big) \leq \sum\limits_{\omega\in[N]} p_\omega \boldsymbol{\xi}_\omega^2 \leq \overline{\tau}_2\big(\boldsymbol{\sigma}_0(\boldsymbol{x})^2 + \boldsymbol{\mu}_0(\boldsymbol{x})^2\big), \\ \sum\limits_{\omega\in[N]} p_\omega = 1 \end{array} \right\}, \tag{7}$$

where $\boldsymbol{\mu}_0(\boldsymbol{x})$ and $\boldsymbol{\sigma}_0(\boldsymbol{x})^2$ denote the vector of (empirical) decision-dependent mean and variance of the random vector $\boldsymbol{\xi}$. Moreover, τ_1, $\underline{\tau}_2$, and $\overline{\tau}_2$ determine the maximum allowable deviations on the empirical first- and second-order moments.

1.2 Contributions

We consider a distributionally robust two-stage stochastic mixed-integer program with continuous recourse, where the distributional ambiguity is modeled with a decision-dependent ambiguity set of probability distributions. We investigate cases where the recourse function and the ambiguity set are either generic or have a special convex/nonconvex structure. Our main goals in this paper are to provide reformulations and finitely-convergent decomposition-based cutting plane algorithms to solve them optimally (or near-optimality). Here is a list of cases and our contributions:

- *Generic recourse function and generic decision-dependent ambiguity set*: We reformulate the problem as a two-stage stochastic mixed-integer nonlinear program (MINLP), where the nonlinearity in the first stage is due to the decision dependency. When the recourse function is convex, for instance, with an objective function in the form of (4), we develop a decomposition-based cutting plane algorithm in virtue of L-shaped algorithm [9, 69] (Section 3.1). When the recourse function is nonconvex, with an objective function in the form of (5), we develop a disjunctive cutting plane algorithm, where the restricted master problem involves solving a noncovex minimization problem (Section 3.2). This algorithm is of independent interest to solve two-stage stochastic programs with a random decision-dependent recourse matrix (i.e., bilinear stochasticity on the left-hand side) or a bilinear objective function (see Remark 6).

- *Convex recourse function and an ambiguity set with either nonnegative convex or nonpositive concave decision-dependency*: We reformulate the problem as a two-stage stochastic mixed-integer concave minimization program. We develop a decomposition-based cutting plane algorithm in virtue of L-shaped algorithm, where the restricted master problem involves solving a concave minimization problem (Section 3.2).

- We provide computational experiments for joint pricing and stocking decisions on a stylized multiproduct newsvendor problem with price-dependent demand involving a nonconvex recourse function. We compare the performance of the proposed decomposition-based cutting plane algorithm with solving the extensive formulation using a commercial nonconvex solver.

1.3 Literature Review

Decision-making problems with endogenous uncertainty are studied in the context of stochastic programming [20, 25], robust optimization [70, 71], and DRO [36, 43], with applications including offshore oilfield exploitation [32], clinical trial planning [17], radiation therapy [41], network design and facility location [1, 5, 44], and power capacity expansion [4]. Following the earlier work of [21], two distinct classes of endogenous uncertainty were identified in [26], depending on whether the decisions affect the *temporal revelation of uncertainty or its probability distribution*—with the possibility of affecting the set of possible outcomes in both classes [42]. Given that our studied stochastic program involves a decision-dependent probability distribution, we limit our literature review to this class of problems.

Unlike models with temporal endogenous uncertainty, the literature on models where decisions affect the probability distribution is sparse. Various discrepancy-based DRO approaches were developed in [36], where to model decision-dependent probability distributions, the maximum allowed distance from a nominal distribution is assumed to be decision-dependent. The authors in [43] considered a DRO approach, where the decision-dependent distributional ambiguity is modeled via Wasserstein distance around a decision-dependent nominal distribution. Similarly, the authors in [4] studied a DRO approach to a facility location problem, where the distributional ambiguity of decision-dependent demand is captured with moments around a decision-dependent nominal moment. A similar problem to the one in [4] was studied in a multistage setting in [74]. The authors in [34, 37] studied a service center location problem, where utility gains upon receiving service are location-dependent and assumed to be ambiguously described by moment-based sets. A joint stocking and pricing problem for a product without knowing the price-dependent demand was studied in [29]. The authors introduced a functionally robust approach to hedge against various classes of decreasing convex or concave functions to model price-dependent demand. A DRO problem with a decision-dependent ambiguity set was studied in [57], where the preference relationship between probability distributions is formed via the decision-dependent cumulative distribution functions.

Similar to the existing models on DRO problems where decisions impact the probability distribution of random parameters, we capture the decision dependency with a decision-dependent ambiguity set. Nevertheless, instead of focusing on a particular type of ambiguity set, e.g., a discrepancy-based or moment-based model, we consider a polyhedral ambiguity set, which subsumes some classes of both models. When the polyhedral ambiguity set (6) subsumes a discrepancy-based model, we allow for a decision-dependent radius, whereas when it subsumes a moment-based model, we allow for decision-dependent nominal parameters.

1.4 Organization

The rest of this paper is outlined as follows. In Section 2, we present reformulations for a general DRO problem with a decision-dependent ambiguity set and special cases of convex/nonconvex recourse functions as well as convex/nonconvex decision-dependency. In Section 3, we propose finitely-convergent decomposition-based cutting plane algorithms to solve the resulting DRO problems optimally (or near-optimally). We then present numerical experiments to test the efficacy of solution algorithms in Section 4. We end with conclusions and a discussion of future work in Section 5.

Notation and Definitions: Throughout this paper, vectors are denoted by boldface lowercase letters and matrices are denoted by boldface uppercase letters. Sets are denoted by calligraphic uppercase letters. All sets in this paper are subsets of a finite-dimensional Euclidean space $\mathbb{R}^d$. For a set $\mathcal{B} \subseteq \mathbb{R}^d$, $\operatorname{conv}(\mathcal{B})$ denote the convex hull of $\mathcal{B}$. Let $\boldsymbol{e}_i$ be the i-th unit vector and $\boldsymbol{e}$ be a vector of ones in $\mathbb{R}^d$. A random function $g : \Xi \mapsto \mathbb{R}$ has N outcomes $\{g(\boldsymbol{\xi}^1), \dots, g(\boldsymbol{\xi}^N)\}$ with probabilities $\{p_1, \dots, p_N\}$. A set-valued function $t \mapsto F(t) : \mathcal{D} \rightrightarrows \mathcal{R}$ is upper semicontinuous (u.s.c.) at $t \in \mathcal{D}$ if $\lim_{v\to\infty} t_v = t$, $y_v \in F(t_v)$, and $\lim_{v\to\infty} y_v = y$ imply that $y \in F(t)$. A real-valued function $t \mapsto F(t) : \mathcal{D} \to \mathbb{R}$ is lower semicontinuous (l.s.c.) at $t \in \mathcal{D}$ if $\lim_{v\to\infty} t_v = t$ implies that $\liminf_{v\to\infty} F(t_v) = F(t)$. For $a \in \mathbb{R}$, $(a)_+$ denotes $\max\{0, a\}$.

2 Reformulations

In this section, we provide reformulations for problem (1) with a decision-dependent polyhedral ambiguity set in the form of (6). In Section 2.1, we focus on a generic recourse function and an ambiguity set with a generic decision dependency. In Section 2.2, we focus on a convex recourse function and an ambiguity set with a convex or concave decision-dependency. We suppose that Assumptions **(A1)**–**(A3)** hold throughput the paper.

2.1 Generic Recourse Function and Ambiguity Set with Generic Decision Dependency

In this section, we suppose that $h(\boldsymbol{x}, \boldsymbol{\xi})$, defined in (2), is a generic function in $\boldsymbol{x}$ (convex or nonconvex) for a fixed $\boldsymbol{\xi} \in \Xi$. Moreover, we do not impose any additional structure on the ambiguity set (6). Theorem 1 states a nonlinear reformulation of (1).

Theorem 1. *Suppose that Assumptions* **(A1)**–**(A3)** *hold. Then,* (1) *can be written as the optimal value of the following nonlinear program:*

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{\lambda}}\ & \boldsymbol{\lambda}^\top \boldsymbol{\vartheta}(\boldsymbol{x}) + \sum_{\omega \in [N]} G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}) \\ \textit{s.t.}\quad & \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{\lambda} \geq \boldsymbol{0}, \end{aligned} \tag{8}$$

where

$$G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}) := \varphi_\omega \Big[h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega) \Big], \tag{9}$$

with

$$\varphi_\omega[z] = \overline{p}_\omega (z)_+ - \underline{p}_\omega (-z)_+. \tag{10}$$

Proof. Let us define $\mathcal{M} := \Big\{ \boldsymbol{p} \geq \boldsymbol{0} \,\Big|\, \underline{p}_\omega \leq p_\omega \leq \overline{p}_\omega,\ \omega \in [N] \Big\}$. For a fixed $\boldsymbol{x} \in \mathcal{X}$, by dualizing the second set of constraints in (6), a Lagrangian function of problem $\max_{\boldsymbol{p} \in \mathcal{P}(\boldsymbol{x})} \ \mathbb{E}_{\boldsymbol{p}} [h(\boldsymbol{x}, \boldsymbol{\xi})]$ can be written as:

$$L(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{\lambda}) = \boldsymbol{\lambda}^\top \boldsymbol{\vartheta}(\boldsymbol{x}) + \sum_{\omega \in [N]} p_\omega \Big(h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega) \Big).$$

Hence, the Lagrangian dual of problem $\max_{\boldsymbol{p} \in \mathcal{P}(\boldsymbol{x})} \ \mathbb{E}_{\boldsymbol{p}} [h(\boldsymbol{x}, \boldsymbol{\xi})]$ is

$$\begin{aligned} \min_{\boldsymbol{\lambda}} \max_{\boldsymbol{p} \in \mathcal{M}}\ & L(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{\lambda}) \\ \text{s.t.}\quad & \boldsymbol{\lambda} \geq \boldsymbol{0}. \end{aligned}$$

Observe that

$$\begin{aligned} & -\boldsymbol{\lambda}^\top \boldsymbol{\vartheta}(\boldsymbol{x}) + \max_{\boldsymbol{p} \in \mathcal{M}} L(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{\lambda}) \\ &= \sum_{\omega \in [N]} \overline{p}_\omega \Big(h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega) \Big)_+ - \sum_{\omega \in [N]} \underline{p}_\omega \Big(-h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) + \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega) \Big)_+ \\ &= \sum_{\omega \in [N]} \overline{p}_\omega \Big(h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega) \Big)_+ - \underline{p}_\omega \Big(-h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) + \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega) \Big)_+ \\ &= \sum_{\omega \in [N]} \varphi_\omega \Big[h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega) \Big] \end{aligned}$$

$$= \sum_{\omega \in [N]} G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}),$$

where $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ and $\varphi_\omega(\cdot)$ are defined in (9) and (10), respectively. We concluded the second equality above by the fact that either $\Big(h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\Big)_+$ or $\Big(-h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) + \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\Big)_+$ is positive. Now, because $\max_{\boldsymbol{p} \in \mathcal{P}(\boldsymbol{x})} \mathbb{E}_{\boldsymbol{p}}[h(\boldsymbol{x}, \boldsymbol{\xi})]$ is a linear program and by Assumption **(A3)**, there is no duality gap. Combining the resulting dual problem with the outer minimization problem derives the reformulation in the statement of the theorem. □

Note that $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$, defined in (9), is well-defined by the finiteness of $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ (see Remark 1). In Lemma 1, we establish that $\varphi_\omega[z]$ is a convex function, and using this lemma, we derive a two-stage reformulation of (1) in Corollary 1.

Lemma 1. *Consider function $\varphi_\omega[z]$, defined in (10), for $\omega \in [N]$. Then, $\varphi_\omega[z]$ is convex and monotonically nondecreasing in z.*

Proof. Observe that $\varphi_\omega[z]$ can rewritten as $\varphi_\omega[z] = \underline{p}_\omega z + \varepsilon(z)_+$, where $\varepsilon \geq 0$ is such that $\overline{p}_\omega = \underline{p}_\omega + \varepsilon$. Hence, $\varphi_\omega[z]$ is monotonically nondecreasing in z. Because $(z)_+$ is convex in z, then, it follows that $\varphi_\omega[z]$ is convex in z. □

Corollary 1. *Suppose that Assumptions* **(A1)**–**(A3)** *hold. Then, (1) can be written as the optimal value of the following two-stage stochastic mixed-integer nonlinear program (MINLP):*

$$\begin{aligned} \min_{\boldsymbol{x}, \boldsymbol{\lambda}} \;& \boldsymbol{\lambda}^\top \boldsymbol{\vartheta}(\boldsymbol{x}) + \sum_{\omega \in [N]} G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}) \\ \text{s.t.} \;& \boldsymbol{x} \in \mathcal{X}, \; \boldsymbol{\lambda} \geq \boldsymbol{0}, \end{aligned} \tag{11}$$

where

$$\begin{aligned} G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}) = \min_{\boldsymbol{y}, \mu, \gamma} \;& \gamma \overline{p}_\omega - \mu \underline{p}_\omega \\ & \gamma - \mu \geq \psi_0(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega), \\ & \boldsymbol{y} \in \mathcal{Y}(\boldsymbol{x}, \boldsymbol{\xi}_\omega), \\ & \gamma, \mu \geq 0. \end{aligned} \tag{12}$$

Proof. By Lemma 1, $\varphi_\omega[z]$, defined in (10), is monotonically nondecreasing in z. Hence,

$$\begin{aligned} G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}) =& \varphi_\omega\Big[h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\Big] \\ =& \min_{\boldsymbol{y} \in \mathcal{Y}(\boldsymbol{x}, \boldsymbol{\xi}_\omega)} \varphi_\omega\Big[\psi_0(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\Big]. \end{aligned}$$

Now, linearization of $\varphi_\omega[\cdot]$ by introducing additional variables γ and μ, yields (12). Theorem 1 completes the proof. □

Remark 2. *It is evident from (11) that once decision variables $(\boldsymbol{x}, \boldsymbol{\lambda})$ are fixed, the problem becomes decomposable in $\omega \in [N]$. In particular, one can interpret (11) as a two-stage stochastic program, where $(\boldsymbol{x}, \boldsymbol{\lambda})$ are the first-stage decisions and $(\boldsymbol{y}_\omega, \mu_\omega, \gamma_\omega)$, $\omega \in [N]$, are the second-stage decisions. In particular, observe that in the case that $\mathcal{P}(\boldsymbol{x})$ is decision-independent, i.e., $\vartheta(\boldsymbol{x})$ is a constant, problem (11) reduces to a two-stage stochastic mixed-integer program.*

We end this section by presenting an extensive deterministic equivalent formulation (DEF) of (11) in Corollary 2.

Corollary 2. *Suppose that Assumptions* **(A1)**–**(A3)** *hold. Then,* (1) *can be written as the optimal value of the following MINLP:*

$$\begin{aligned}\min_{\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{y}_1,\dots,\boldsymbol{y}_N,\boldsymbol{\gamma},\boldsymbol{\mu}} \quad & \boldsymbol{\lambda}^\top \boldsymbol{\vartheta}(\boldsymbol{x}) + \sum_{\omega\in[N]} (\gamma_\omega \overline{p}_\omega - \mu_\omega \underline{p}_\omega) \\ s.t. \quad & \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{\lambda} \geq \mathbf{0}, \\ & \gamma_\omega - \mu_\omega \geq \psi_0(\boldsymbol{x}, \boldsymbol{y}_\omega, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega), \quad \omega \in [N], \\ & \boldsymbol{y}_\omega \in \mathcal{Y}(\boldsymbol{x}, \boldsymbol{\xi}_\omega), \quad \omega \in [N], \\ & \boldsymbol{\gamma}, \boldsymbol{\mu} \geq \mathbf{0}.\end{aligned} \tag{13}$$

2.2 Convex Recourse Function and Ambiguity Set with Convex/Concave Decision Dependency

Throughout this section, we suppose that $h(\boldsymbol{x}, \boldsymbol{\xi})$, defined in (2), is a convex function in $\boldsymbol{x}$ for a fixed $\boldsymbol{\xi} \in \Xi$; for instance $\psi_0(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\xi}) = \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{q}^\top \boldsymbol{y}$ as in (4). Proposition 1 states that in this case, $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$, defined in (9), is convex in $(\boldsymbol{x}, \boldsymbol{\lambda})$. Then, Corollary 3 below states that under certain conditions on the ambiguity set $\mathcal{P}(\boldsymbol{x})$, problem (11) can be stated as a two-stage stochastic mixed-integer concave minimization problem.

Proposition 1. *Suppose that Assumptions* **(A1)**–**(A3)** *hold. If $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ is convex in $\boldsymbol{x}$ for $\omega \in [N]$, then $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$, defined in* (9)*, is convex in $(\boldsymbol{x}, \boldsymbol{\lambda})$ for $\omega \in [N]$.*

Proof. Observe that $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)$ is convex in $(\boldsymbol{x}, \boldsymbol{\lambda})$. Hence, $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ is convex in $(\boldsymbol{x}, \boldsymbol{\lambda})$ because $\varphi_\omega(\cdot)$ preserves convexity. □

To state the main result in this section, we make the following assumption on the ambiguity set.

A4 (Convex/Concave Decision-Dependent Ambiguity Set) For the ambiguity set $\mathcal{P}(\boldsymbol{x})$, defined in (6), $\vartheta^i(\boldsymbol{x})$, is either a nonnegative convex function in $\boldsymbol{x}$ or a nonpositive concave function in $\boldsymbol{x}$, for $i \in [s]$.

Corollary 3. *Suppose that Assumptions* **(A1)**–**(A4)** *hold. For $\omega \in [N]$, suppose that $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ is a convex recourse function defined in* (2)*. Then,* (1) *can be written as the optimal value of the following two-stage stochastic mixed-integer concave minimization problem:*

$$\begin{aligned}\min_{\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\theta},\Psi} \quad & \Psi + \sum_{\omega\in[N]} \theta_\omega - \frac{1}{2}\sum_{i\in\mathcal{C}^+} \left(\lambda_i^2 + \vartheta^i(\boldsymbol{x})^2\right) - \frac{1}{2}\sum_{i\in\mathcal{C}^-} \left(\lambda_i - \vartheta^i(\boldsymbol{x})\right)^2 \\ s.t. \quad & \theta_\omega \geq G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}), \quad \omega \in [N], \\ & \Psi \geq \frac{1}{2}\sum_{i\in\mathcal{C}^+} \left(\lambda_i + \vartheta^i(\boldsymbol{x})\right)^2 + \frac{1}{2}\sum_{i\in\mathcal{C}^-} \left(\lambda_i^2 + \vartheta^i(\boldsymbol{x})^2\right) \\ & \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{\lambda} \geq \mathbf{0},\end{aligned} \tag{14}$$

where $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ is defined as in (12)*, $\mathcal{C}^+ := \left\{i \in [s] \,\middle|\, \vartheta^i(\boldsymbol{x}) \geq 0 \text{ and convex in } \boldsymbol{x}\right\}$, and $\mathcal{C}^- := \left\{i \in [s] \,\middle|\, \vartheta^i(\boldsymbol{x}) \leq 0 \text{ and concave in } \boldsymbol{x}\right\}$.*

Proof. Note that $2ab = (a+b)^2 - (a^2+b^2)$. Using this identity, we can write $2\lambda_i\vartheta^i(\boldsymbol{x})$ as the difference of two convex functions as follows $2\lambda_i\vartheta^i(\boldsymbol{x}) = \left(\lambda_i + \vartheta^i(\boldsymbol{x})\right)^2 - \left(\lambda_i^2 + \vartheta^i(\boldsymbol{x})^2\right)$ for $i \in \mathcal{C}^+$. Similarly, we can write $2\lambda_i\vartheta^i(\boldsymbol{x}) = \big(\lambda_i^2 + \vartheta^i(\boldsymbol{x})^2\big) - \big(\lambda_i - \vartheta^i(\boldsymbol{x})\big)^2$ for $i \in \mathcal{C}^-$, which is again the difference of two convex functions. Problem (14) is then obtained from (11), where the concave terms are kept in the objective function and the convex terms are moved to the constraints by introducing additional variables θ and Ψ. Because the objective function is concave and the constraints induce a convex feasible region, the problem is a two-stage stochastic mixed-integer concave minimization problem. □

3 Solution Algorithms

The two-stage structure makes problem (11) amenable to decomposition-based algorithms. This section investigates decomposition-based cutting plane approaches to obtain an optimal (or near-optimal) solution to (11) under different assumptions on the recourse function and the decision-dependent ambiguity set. These cases will involve solving a nonconvex problem; hence, we assume that there is an oracle that solves this problem to ϵ-optimality as follows.

A5 (ϵ-Optimal Global Solver) For any $\epsilon > 0$, there is an oracle that solves a nonconvex problem $z^* = \min_{(\boldsymbol{x},\boldsymbol{\lambda},\theta)\in\mathcal{S}} \boldsymbol{\lambda}^\top \boldsymbol{\vartheta}(\boldsymbol{x}) + \theta$ to ϵ-optimality. That is, it obtains a solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}, \hat{\theta})$ such that $\boldsymbol{\vartheta}(\hat{\boldsymbol{x}})^\top \hat{\boldsymbol{\lambda}} + \hat{\theta} \leq z^* + \epsilon$.

3.1 Convex Recourse Function and Ambiguity Set with Generic Decision Dependency

Throughout this section, we suppose that $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ is a convex function in $\boldsymbol{x}$, $\omega \in [N]$. Before presenting an algorithm for solving (1) with a generic decision-dependent ambiguity set (6), we review such an algorithm for a decision-independent ambiguity set.

Recall that when $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ is a convex function in $\boldsymbol{x}$, $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$, defined in (9), is convex in $(\boldsymbol{x}, \boldsymbol{\lambda})$, $\omega \in [N]$ (Proposition 1). In addition, when the ambiguity set $\mathcal{P}(\boldsymbol{x})$ is decision-independent, i.e., $\vartheta^i(\boldsymbol{x}) = \vartheta^i$, $i \in [s]$, (11) reduces to a convex program. Consequently, by exploiting the convex structure of $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$, one can iteratively obtain outer approximations to $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ for $\omega \in [N]$ using subgradient information. These, in turn, lead to a lower bound on the optimal value. Below, we first present subdifferential of $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$, $\omega \in [N]$. Then, we discuss the general framework of such a cutting plane scheme.

Proposition 2. *Suppose that Assumptions* **(A1)**–**(A3)** *hold. For $\omega \in [N]$, suppose that $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ is a convex recourse function defined in (2). Then, for $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$, defined in (9), we have*

$$\partial G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}) = \begin{cases} \overline{p}_\omega \partial h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) \times \{-\boldsymbol{g}(\boldsymbol{\xi}_\omega)\} & \text{if } h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega) > 0, \\ \underline{p}_\omega \partial h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) \times \{-\boldsymbol{g}(\boldsymbol{\xi}_\omega)\} & \text{if } h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega) < 0, \\ \bigcup_{\eta \in [\underline{p}_\omega, \overline{p}_\omega]} \eta \partial h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) \times \{-\boldsymbol{g}(\boldsymbol{\xi}_\omega)\} & \text{if } h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega) = 0, \end{cases} \tag{15}$$

where $\times$ denotes the Cartesian product.

Proof. Observe that by Proposition 1, $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ is convex in $(\boldsymbol{x}, \boldsymbol{\lambda})$. Thus, the subdifferential (15) follows immediately from (9). □

Given $\hat{\boldsymbol{x}} \in \mathcal{X}$ and $\hat{\boldsymbol{\lambda}} \geq \boldsymbol{0}$, observe that by using the subgradient inequality, we have

$$G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}) \geq G_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) + \boldsymbol{\zeta}_{\boldsymbol{x},\omega}^\top (\boldsymbol{x} - \hat{\boldsymbol{x}}) + \boldsymbol{\zeta}_{\boldsymbol{\lambda},\omega}^\top (\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}), \tag{16}$$

where $\boldsymbol{\zeta}_\omega := (\boldsymbol{\zeta}_{\boldsymbol{x},\omega}^\top, \boldsymbol{\zeta}_{\boldsymbol{\lambda},\omega}^\top)^\top \in \partial G_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}})$. Hence, for the epigraph

$$\mathfrak{Q}_\omega := \{(\boldsymbol{x}, \boldsymbol{\lambda}, \theta_\omega) \,|\, G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}) \le \theta_\omega,\ \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{\lambda} \ge \mathbf{0}\},$$

we can obtain supporting hyperplanes to separate point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}})$ from the epigraph. The supporting hyperplanes are of the form

$$\theta_\omega \ge G_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) + \boldsymbol{\zeta}_{\boldsymbol{x},\omega}^\top(\boldsymbol{x} - \hat{\boldsymbol{x}}) + \boldsymbol{\zeta}_{\boldsymbol{\lambda},\omega}^\top(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}). \tag{17}$$

Therefore, one can obtain a restricted master problem for (1) after t iterations as follows:

$$z^t = \min_{(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\theta}) \in \mathcal{S}^t} \boldsymbol{\lambda}^\top \boldsymbol{\vartheta} + \sum_{\omega \in [N]} \theta_\omega, \tag{18}$$

where

$$\mathcal{S}^t = \left\{ (\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\theta}) \;\middle|\; \begin{array}{l} \boldsymbol{A}\boldsymbol{x} \ge \boldsymbol{d},\ \boldsymbol{x} \ge \mathbf{0},\ \boldsymbol{\lambda} \ge \mathbf{0}, \\ \theta_\omega \ge G_\omega(\boldsymbol{x}^k, \boldsymbol{\lambda}^k) + (\boldsymbol{x} - \boldsymbol{x}^k)^\top \boldsymbol{\zeta}_{\boldsymbol{x},\omega}^k + \\ \quad (\boldsymbol{\lambda} - \boldsymbol{\lambda}^k)^\top \boldsymbol{\zeta}_{\boldsymbol{\lambda},\omega}^k,\ \omega \in [N],\ k \in \{0\} \cup [t-1] \end{array} \right\}, \tag{19}$$

$(\boldsymbol{x}^t, \boldsymbol{\lambda}^t, \boldsymbol{\theta}^t)$ is a solution obtained from solving (18) at iteration t, and $\boldsymbol{\zeta}^k$ are the corresponding subgradient of $G_\omega(\boldsymbol{x}^k, \boldsymbol{\lambda}^k)$, $\omega \in [N]$ and $k \in \{0\} \cup [t-1]$. Note that $\mathcal{S}^0$ is the initial set, with no optimality cuts.

It is clear that $\sum_{\omega \in [N]} Q_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ is also convex in $(\boldsymbol{x}, \boldsymbol{\lambda})$. Hence, one can also obtain a lower approximation to (8) where one generates cuts of the form $\theta \ge \sum_{\omega \in [N]} \Big(G_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) + \boldsymbol{\zeta}_{\boldsymbol{x},\omega}^\top(\boldsymbol{x} - \hat{\boldsymbol{x}}) + \boldsymbol{\zeta}_{\boldsymbol{\lambda},\omega}^\top(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}) \Big)$.

When $h(\boldsymbol{x}, \boldsymbol{\xi})$ is defined with objective function $\psi_0(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\xi}) = \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{q}^\top \boldsymbol{y}$, as in (4), we have that $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ as the optimal value of a linear program as follows:

$$G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}) = \min_{\boldsymbol{y},\mu,\gamma} \gamma \overline{p}_\omega - \mu \underline{p}_\omega \tag{20a}$$

$$\gamma - \mu - \boldsymbol{q}_\omega^\top \boldsymbol{y} \ge \boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega), \tag{20b}$$

$$\boldsymbol{D}_\omega \boldsymbol{y} \ge \boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega, \tag{20c}$$

$$\boldsymbol{y} \ge \mathbf{0},\ \gamma, \mu \ge 0. \tag{20d}$$

By taking the dual of (20), we have

$$\begin{aligned} G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}) = \max_{\boldsymbol{\pi},\sigma}\ & \sigma\big(\boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\big) + \boldsymbol{\pi}^\top(\boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega) \\ \text{s.t.}\ \ & \underline{p}_\omega \le \sigma \le \overline{p}_\omega, \\ & -\sigma \boldsymbol{q}_\omega^\top + \boldsymbol{\pi}^\top \boldsymbol{D}_\omega \le \mathbf{0}^\top, \\ & \boldsymbol{\pi} \ge \mathbf{0},\ \sigma \ge 0. \end{aligned} \tag{21}$$

where σ and $\boldsymbol{\pi}$ are the dual vectors corresponding to (20b) and (20c), respectively. Let $\big(\hat{\sigma}_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}), \hat{\pi}_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}})\big)$ be an optimal dual multiplier to problem (21) at $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}})$. Then, an optimality cut in the form of (17) can be written as:

$$\theta_\omega \ge \hat{\sigma}_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}})\big(\boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\big) + \hat{\boldsymbol{\pi}}_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}})^\top(\boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega). \tag{22}$$

For simplicity, we let

$$\begin{aligned} &\boldsymbol{\alpha}_{\boldsymbol{x},\omega}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) := \boldsymbol{B}_\omega^\top \hat{\boldsymbol{\pi}}_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) + \boldsymbol{c}_\omega \hat{\sigma}_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) \\ &\boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) := -\boldsymbol{g}(\boldsymbol{\xi}_\omega) \hat{\sigma}_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) \\ &\rho_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) := \boldsymbol{b}_\omega^\top \hat{\boldsymbol{\pi}}_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}). \end{aligned} \tag{23}$$

Algorithm 1: Cutting-plane algorithm for problem (1) with a convex recourse function and a decision-independent ambiguity set.

Input: An initial solution $(\boldsymbol{x}^0, \boldsymbol{\lambda}^0)$.
Output: An optimal solution and the optimal value.

Initialization: Set $t \leftarrow 1$, $\text{LB} \leftarrow -\infty$, $\text{UB} \leftarrow +\infty$. Add initial cuts (e.g., $\theta_\omega \geq 0$), if available, to $\mathcal{S}^0$.
while UB > LB **do**
  for *each* $\omega \in [N]$ **do**
    Obtain $G_\omega(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1})$ by solving (12) and subgradients $\boldsymbol{\zeta}_\omega \in \partial G_\omega(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1})$ using (15).
  end
  $\text{UB} \leftarrow \min\{\text{UB}, \boldsymbol{\vartheta}^\top \boldsymbol{\lambda}^{t-1} + \sum_{\omega \in [N]} G_\omega(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1})\}$.
  Let $\mathcal{S}^t = \mathcal{S}^{t-1} \cap \big\{(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\theta}) \,\big|\, \theta_\omega \geq G_\omega(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1}) + \boldsymbol{\zeta}_{\boldsymbol{x},\omega}^\top(\boldsymbol{x} - \boldsymbol{x}^{t-1}) + \boldsymbol{\zeta}_{\boldsymbol{\lambda},\omega}^\top(\boldsymbol{\lambda} - \boldsymbol{\lambda}^{t-1}),\ \omega \in [N]\big\}$.
  Solve restricted master problem (18) using $\mathcal{S}^t$ and obtain an optimal solution $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t, \boldsymbol{\theta}^t)$.
  Let $\text{LB} \leftarrow \boldsymbol{\vartheta}^\top \boldsymbol{\lambda}^t + \sum_{\omega \in [N]} \theta_\omega^t$.
  Set $t \leftarrow t + 1$.
end
return $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$ and UB.

Hence,

$$\theta_\omega \geq \boldsymbol{x}^\top \boldsymbol{\alpha}_{\boldsymbol{x},\omega}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) + \boldsymbol{\lambda}^\top \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) + \rho_\omega(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}) \tag{24}$$

is a valid inequality in the form of (17).

A general framework of a finitely-convergent cutting plane scheme to solve (11) with a convex recourse function and a decision-independent ambiguity set is given in Algorithm 1.

Theorem 2. *Suppose that Assumptions* **(A1)**–**(A3)** *hold. For $\omega \in [N]$, suppose that $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ is a convex recourse function defined in* (2)*. Moreover, $\mathcal{P}(\boldsymbol{x})$, defined in* (6)*, is decision-independent. Then, Algorithm 1 generates an optimal solution to* (1) *in a finite number of iterations.*

Proof. To prove the finite convergence of Algorithm 1, we need to show that the "while" loop terminates in a finite number of iterations, generating an optimal solution to (1).

First, note that at each iteration t, $G_\omega(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1})$ is finite, $\omega \in [N]$. Now, after solving (18) using $\mathcal{S}^t$ and obtaining an optimal solution $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t, \boldsymbol{\theta}^t)$, either (1) $\theta_\omega^t \geq G_\omega(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$ for all $\omega \in [N]$ or (2) $\theta_\omega^t < G_\omega(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$ for some $\omega \in [N]$. If case (1) happens, we have

$$\text{LB} = \boldsymbol{\vartheta}^\top \boldsymbol{\lambda}^t + \sum_{\omega \in [N]} \theta_\omega^t \geq \boldsymbol{\vartheta}^\top \boldsymbol{\lambda}^t + \sum_{\omega \in [N]} G_\omega(\boldsymbol{x}^t, \lambda^t) = \text{UB}.$$

Hence, the termination criterion is satisfied, and Algorithm 1 returns the optimal value UB, with the corresponding optimal solution $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$.

Now, suppose that case (2) happens. It can be verified that solution $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t, \boldsymbol{\theta}^t)$ is not feasible to the epigraphic reformulation of problem (1) for some $\omega \in [N]$, because $\theta_\omega^t < G_\omega(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$. Also, we have

$$\text{LB} = \boldsymbol{\vartheta}^\top \boldsymbol{\lambda}^t + \sum_{\omega \in [N]} \theta_\omega^t < \boldsymbol{\vartheta}^\top \boldsymbol{\lambda}^t + \sum_{\omega \in [N]} G_\omega(\boldsymbol{x}^t, \lambda^t) = \text{UB}.$$

Because the termination criterion is not satisfied, we derive optimality cut $\theta_\omega \geq G_\omega(\boldsymbol{x}^t, \boldsymbol{\lambda}^t) + \boldsymbol{\zeta}_{\boldsymbol{x},\omega}^\top(\boldsymbol{x}-\boldsymbol{x}^t)+\boldsymbol{\zeta}_{\boldsymbol{\lambda},\omega}^\top(\boldsymbol{\lambda}-\boldsymbol{\lambda}^t)$. Now, we show that case (2) happens a finite number of iterations. Note that $\theta_\omega^t < G_\omega(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$ implies that none of the previously generated cuts enforce $\theta_\omega \geq G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ at $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$. Therefore, a new subgradient $\boldsymbol{\zeta}_\omega = (\boldsymbol{\zeta}_{\boldsymbol{x},\omega}^\top, \boldsymbol{\zeta}_{\boldsymbol{\lambda},\omega}^\top)^\top$ of function $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ at $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$ is needed to cut off the point $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t, \boldsymbol{\theta}^t)$. Observe that $\boldsymbol{\zeta}_{\boldsymbol{x},\omega}$ corresponds to one of the finitely many dual bases of $h(\boldsymbol{x}^t, \boldsymbol{\xi}_\omega)$; that is, $\left\{\boldsymbol{\pi} \geq \mathbf{0} \,\middle|\, \boldsymbol{\pi}^\top \boldsymbol{D}_\omega \leq \boldsymbol{q_\omega}^\top\right\}$. Hence, there are finitely many optimality cuts. Consequently, case (2) can only happen a finite number of times. This completes the proof. □

Remark 3. *Algorithm 1 can be used for the cases where there are integer variables in the first stage, i.e., $n_1 > 0$, and the finite-convergence analysis remains valid where one solves the corresponding mixed-integer master problem to optimality at each iteration of Algorithm 1. Alternatively, one can use a branch-and-cut procedure, where a linear relaxation of the master problem is solved with the standard branch-and-bound procedure. Then, a globally valid optimality cut is obtained whenever an integer feasible solution to the restricted master problem violates any of the previously generated optimality cuts. By the finiteness of the branch-and-bound, it is straightforward to prove the finite convergence of this adjusted version of Algorithm 1.*

When the ambiguity set $\mathcal{P}(\boldsymbol{x})$ is decision-dependent, similar to the construction for the decision-independent case, one can form a restricted master problem for (1) after t iterations as follows:

$$z^t = \min_{(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\theta})\in\mathcal{S}^t} \boldsymbol{\lambda}^\top \boldsymbol{\vartheta}(\boldsymbol{x}) + \sum_{\omega\in[N]} \theta_\omega, \tag{25}$$

where

$$\mathcal{S}^t = \left\{ (\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\theta}) \,\middle|\, \begin{array}{l} \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{d},\ \boldsymbol{x} \geq \mathbf{0},\ \boldsymbol{\lambda} \geq \mathbf{0}, \\ \theta_\omega \geq G_\omega(\boldsymbol{x}^k, \boldsymbol{\lambda}^k) + (\boldsymbol{x} - \boldsymbol{x}^k)^\top \boldsymbol{\zeta}_{\boldsymbol{x},\omega}^k + \\ \quad (\boldsymbol{\lambda} - \boldsymbol{\lambda}^k)^\top \boldsymbol{\zeta}_{\boldsymbol{\lambda},\omega}^k,\ \omega \in [N],\ k \in \{0\} \cup [t-1] \end{array} \right\}, \tag{26}$$

$(\boldsymbol{x}^t, \boldsymbol{\lambda}^t, \boldsymbol{\theta}^t)$ is a solution obtained from solving (25) at iteration t, and $\boldsymbol{\zeta}^k$ are the corresponding subgradient of $G_\omega(\boldsymbol{x}^k, \boldsymbol{\lambda}^k)$, $\omega \in [N]$ and $k \in \{0\}\cup[t-1]$. A general framework of a finitely-convergent cutting plane scheme to solve (1) with a convex recourse function and a generic decision-dependent ambiguity set is given in Algorithm 2.

Theorem 3. *Suppose that Assumptions* **(A1)**–**(A3)** *and* **(A5)** *hold. For $\omega \in [N]$, suppose that $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ is a convex recourse function defined in (2). Then, Algorithm 2 generates an ϵ-optimal solution to (1) in a finite number of iterations.*

Proof. Using a similar argument as that in the proof of Theorem 2, after a finite number of iterations t, we have $\theta_\omega^t \geq G_\omega(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$ for all $\omega \in [N]$. This implies that

$$\text{LB} = \boldsymbol{\vartheta}(\boldsymbol{x}^t)^\top \boldsymbol{\lambda}^t + \sum_{\omega\in[N]} \theta_\omega^t \geq \boldsymbol{\vartheta}(\boldsymbol{x}^t)^\top \boldsymbol{\lambda}^t + \sum_{\omega\in[N]} G_\omega(\boldsymbol{x}^t, \lambda^t) = \text{UB}.$$

Hence, the termination criterion is satisfied. On the other hand, given Assumption **(A5)**, we have $\text{LB} \leq z^t + \epsilon$. Thus, Algorithm 2 returns the ϵ-optimal value UB, because $\text{UB} \leq z^t + \epsilon$, with the corresponding ϵ-optimal solution $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$. □

Remark 4. *A challenge in obtaining the lower bound at each iteration of Algorithm 2 is the nonlinear product term $\boldsymbol{\lambda}^\top \boldsymbol{\vartheta}(\boldsymbol{x})$. Various techniques are developed for the global optimization and relaxation of nonconvex nonlinear programs, see, e.g., [13, 18, 53, 55, 60, 66, 67] for a finitely-convergent cutting plane algorithm to obtain an ϵ-optimal solution to nonconvex bilinear programs.*

Algorithm 2: Cutting-plane algorithm for problem (1) with a convex recourse function and a generic decision-dependent ambiguity set.

Input: An initial solution $(\boldsymbol{x}^0, \boldsymbol{\lambda}^0)$ and $\epsilon > 0$ for the optimality tolerance.
Output: An ϵ-optimal solution and the ϵ-optimal value.

```
Initialization: Set t ← 1, LB ← −∞, UB ← +∞. Add initial cuts (e.g., θ_ω ≥ 0), if available, to S^0.
while UB > LB do
    for each ω ∈ [N] do
        Obtain G_ω(x^{t−1}, λ^{t−1}) by solving (12) and subgradients ζ_ω ∈ ∂G_ω(x^{t−1}, λ^{t−1}) using (15).
    end
    UB ← min{UB, ϑ(x^{t−1})^⊤ λ^{t−1} + Σ_{ω∈[N]} G_ω(x^{t−1}, λ^{t−1})}.
    Let S^t = S^{t−1} ∩ {(x, λ, θ) | θ_ω ≥ G_ω(x^{t−1}, λ^{t−1}) + ζ_{x,ω}^⊤(x − x^{t−1}) + ζ_{λ,ω}^⊤(λ − λ^{t−1}), ω ∈ [N]}.
    Solve restricted master problem (25) using S^t and obtain an ε-optimal solution (x^t, λ^t, θ^t).
    Let LB ← ϑ(x^t)^⊤ λ^t + Σ_{ω∈[N]} θ_ω^t.
    Set t ← t + 1.
end
return (x^t, λ^t) and UB.
```

These cutting plane-based algorithms have found applications in several branch-and-bound schemes, see, e.g., [66] for a reformulation-linearization technique (RLT)-based branch-and-bound algorithm, and [58, 68] for a spatial branch-and-bound algorithm. For a comprehensive review of such global optimization techniques, interested readers are referred to [16]. In the context of DRO with a decision-dependent ambiguity set, the authors in [4, 74] used McCormick inequalities to obtain an exact reformulation of bilinear terms due to the multiplication of a binary and a continuous variable. Given that many of these global optimization and relaxation techniques are implemented in some commercial nonconvex solvers, we directly leverage them in our numerical experiments in Section 4,

3.2 Convex Recourse Function and Ambiguity Set with Convex/Concave Decision Dependency

We now turn our attention to problem (14), where $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ is a convex function in $\boldsymbol{x}$, $\omega \in [N]$, and Assumption **(A4)** holds for the decision-dependent ambiguity set.

By a similar argument as in Section 3.1, one can obtain lower bounding approximations to (14). To solve this problem, a restricted master problem for (1) after t iterations is as follows:

$$z^t = \min_{(\boldsymbol{x}, \boldsymbol{\lambda}, \theta, \Psi) \in \mathcal{S}^t} \Psi + \sum_{\omega \in [N]} \theta_\omega - \frac{1}{2} \sum_{i \in \mathcal{C}^+} \left(\lambda_i^2 + \vartheta^i(\boldsymbol{x})^2\right) - \frac{1}{2} \sum_{i \in \mathcal{C}^-} \left(\lambda_i - \vartheta^i(\boldsymbol{x})\right)^2, \tag{27}$$

where

$$\mathcal{S}^t = \left\{ (\boldsymbol{x}, \boldsymbol{\lambda}, \theta, \Psi) \;\middle|\; \begin{array}{l} \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{d},\ \boldsymbol{x} \geq \boldsymbol{0},\ \boldsymbol{\lambda} \geq \boldsymbol{0}, \\ \Psi \geq \frac{1}{2} \sum_{i \in \mathcal{C}^+} \left(\lambda_i + \vartheta^i(\boldsymbol{x})\right)^2 + \frac{1}{2} \sum_{i \in \mathcal{C}^-} \left(\lambda_i^2 + \vartheta^i(\boldsymbol{x})^2\right), \\ \theta_\omega \geq Q_\omega(\boldsymbol{x}^k, \boldsymbol{\lambda}^k) + (\boldsymbol{x} - \boldsymbol{x}^k)^\top \boldsymbol{\zeta}_{\boldsymbol{x},\omega}^k + \\ \quad (\boldsymbol{\lambda} - \boldsymbol{\lambda}^k)^\top \boldsymbol{\zeta}_{\boldsymbol{\lambda},\omega}^k,\ \omega \in [N],\ k \in \{0\} \cup [t-1] \end{array} \right\}, \tag{28}$$

Algorithm 3: Cutting-plane algorithm for problem (1) with a convex recourse function and a convex/concave decision-dependent ambiguity set.

Input: An initial solution $(\boldsymbol{x}^0, \boldsymbol{\lambda}^0)$ and $\epsilon > 0$ for the optimality tolerance.
Output: An ϵ-optimal solution and the ϵ-optimal value.

Initialization: Set $t \leftarrow 1$, $\text{LB} \leftarrow -\infty$, $\text{UB} \leftarrow +\infty$. Add initial cuts (e.g., $\theta_\omega \geq 0$), if available, to $\mathcal{S}^0$.
while UB > LB **do**
  for ***each*** $\omega \in [N]$ **do**
    Obtain $G_\omega(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1})$ by solving (12) and subgradients $\boldsymbol{\zeta}_\omega \in \partial G_\omega(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1})$ using (15).
  end
  $\text{UB} \leftarrow \min\{\text{UB}, \Psi^{t-1} + \sum_{\omega\in[N]} G_\omega(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1}) - \frac{1}{2}\sum_{i\in\mathcal{C}^+}\left((\lambda_i^{t-1})^2 + (\vartheta^i(\boldsymbol{x}^{t-1}))^2\right) - \frac{1}{2}\sum_{i\in\mathcal{C}^-}\left(\lambda_i^{t-1} - \vartheta^i(\boldsymbol{x}^{t-1})\right)^2\}$.
  Let $\mathcal{S}^t = \mathcal{S}^{t-1} \cap \left\{(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\theta}) \,\middle|\, \theta_\omega \geq G_\omega(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1}) + \boldsymbol{\zeta}_{\boldsymbol{x},\omega}^\top(\boldsymbol{x} - \boldsymbol{x}^{t-1}) + \boldsymbol{\zeta}_{\boldsymbol{\lambda},\omega}^\top(\boldsymbol{\lambda} - \boldsymbol{\lambda}^{t-1}),\ \omega \in [N]\right\}$.
  Solve restricted master problem (27) using $\mathcal{S}^t$ and obtain an ϵ-optimal solution $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t, \boldsymbol{\theta}^t, \Psi^t)$.
  Let $\text{LB} \leftarrow \Psi^t + \sum_{\omega\in[N]}\theta_\omega^t - \frac{1}{2}\sum_{i\in\mathcal{C}^+}\left((\lambda_i^t)^2 + (\vartheta^i(\boldsymbol{x}^t))^2\right) - \frac{1}{2}\sum_{i\in\mathcal{C}^-}\left(\lambda_i^t - \vartheta^i(\boldsymbol{x}^t)\right)^2$.
  Set $t \leftarrow t + 1$.
end
return $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$ and UB.

$(\boldsymbol{x}^t, \boldsymbol{\lambda}^t, \boldsymbol{\theta}^t, \Psi^t)$ is an (extreme point) solution obtained from solving (27) at iteration t, and $\boldsymbol{\zeta}^k$ are the corresponding subgradient of $G_\omega(\boldsymbol{x}^k, \boldsymbol{\lambda}^k)$, $\omega \in [N]$ and $k \in \{0\} \cup [t-1]$. A general framework of a finitely-convergent cutting plane scheme to solve (14) is given in Algorithm 3.

Theorem 4. *Suppose that Assumptions* **(A1)**–**(A5)** *hold. For $\omega \in [N]$, suppose that $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ is a convex recourse function defined in* (2). *Then, Algorithm 3 generates an ϵ-optimal solution to* (1) *in a finite number of iterations.*

Proof. Using a similar argument as that in the proof of Theorem 2, after a finite number of iterations t, we have $\theta_\omega^t \geq G_\omega(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$ for all $\omega \in [N]$. This implies that

$$\begin{aligned}
\text{LB} &= \Psi^t + \sum_{\omega\in[N]}\theta_\omega^t - \frac{1}{2}\sum_{i\in\mathcal{C}^+}\left((\lambda_i^t)^2 + (\vartheta^i(\boldsymbol{x}^t))^2\right) - \frac{1}{2}\sum_{i\in\mathcal{C}^-}\left(\lambda_i^t - \vartheta^i(\boldsymbol{x}^t)\right)^2 \\
&\geq \Psi^t + \sum_{\omega\in[N]} G_\omega(\boldsymbol{x}^t, \boldsymbol{\lambda}^t) - \frac{1}{2}\sum_{i\in\mathcal{C}^+}\left((\lambda_i^t)^2 + (\vartheta^i(\boldsymbol{x}^t))^2\right) - \frac{1}{2}\sum_{i\in\mathcal{C}^-}\left(\lambda_i^t - \vartheta^i(\boldsymbol{x}^t)\right)^2 \\
&= \text{UB}.
\end{aligned}$$

Hence, the termination criterion is satisfied. On the other hand, given Assumption **(A5)**, we have $\text{LB} \leq z^t + \epsilon$. Thus, Algorithm 3 returns the ϵ-optimal value UB, because $\text{UB} \leq z^t + \epsilon$, with the corresponding ϵ-optimal solution $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$. □

Remark 5. *A challenge in obtaining the lower bound at each iteration of Algorithm 3 is to solve a concave minimization problem* (27). *Given that solution $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t, \boldsymbol{\theta}^t, \Psi^t)$ obtained from solving* (27) *at iteration t is an extreme point, a cutting plane algorithm can be developed, see, e.g., [47, 48]. We also refer to [10, 27] for a review of the global optimization of concave minimization problems.*

3.3 Nonconvex Recourse Function and Ambiguity Set with Generic Decision Dependency

Throughout this section, we suppose that $h(\boldsymbol{x},\boldsymbol{\xi}_\omega)$ is a nonconvex function in $\boldsymbol{x}$, $\omega \in [N]$, unless otherwise stated. In particular, we assume that $h(\boldsymbol{x},\boldsymbol{\xi}_\omega)$ is defined with objective function $\psi_0(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\xi}_\omega) = \boldsymbol{c}_\omega^\top \boldsymbol{x} + \boldsymbol{q}_\omega^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{L}_\omega \boldsymbol{y}$, as in (5).

Observe that $G_\omega(\boldsymbol{x},\boldsymbol{\lambda})$, defined in (12), is the optimal value of a linear program as follows:

$$
\begin{aligned}
G_\omega(\boldsymbol{x},\boldsymbol{\lambda}) = \min_{\boldsymbol{y},\mu,\gamma}\ & \gamma\overline{p}_\omega - \mu\underline{p}_\omega && \text{(29a)}\\
& \gamma - \mu - \boldsymbol{q}_\omega^\top \boldsymbol{y} - \boldsymbol{x}^\top \boldsymbol{L}_\omega \boldsymbol{y} \geq \boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega), && \text{(29b)}\\
& \boldsymbol{D}_\omega \boldsymbol{y} \geq \boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega, && \text{(29c)}\\
& \boldsymbol{y} \geq \boldsymbol{0},\ \gamma,\mu \geq 0. && \text{(29d)}
\end{aligned}
$$

Let σ and $\boldsymbol{\pi}$ be the Lagrange multipliers corresponding to (29b) and (29c), respectively. Moreover, $\boldsymbol{\pi}_{\boldsymbol{y}}$, π_γ, and π_μ are the Lagrange multipliers corresponding to $\boldsymbol{y} \geq \boldsymbol{0}$, $\gamma \geq 0$, and $\mu \geq 0$. The Lagrangian function can be written as:

$$
\begin{aligned}
& L_\omega(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{y},\mu,\gamma,\sigma,\boldsymbol{\pi},\boldsymbol{\pi}_{\boldsymbol{y}},\pi_\gamma,\pi_\mu)\\
&= \gamma\overline{p}_\omega - \mu\underline{p}_\omega + \sigma(\boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega) - \gamma + \mu + \boldsymbol{q}_\omega^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{L}_\omega \boldsymbol{y}) +\\
&\quad \boldsymbol{\pi}^\top(\boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega - \boldsymbol{D}_\omega \boldsymbol{y}) - \boldsymbol{\pi}_{\boldsymbol{y}}^\top \boldsymbol{y} - \pi_\gamma\gamma - \pi_\mu\mu\\
&= \sigma\big(\boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\big) + \boldsymbol{\pi}^\top(\boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega) + \gamma(\overline{p}_\omega - \pi_\gamma - \sigma) - \mu(\underline{p}_\omega + \pi_\mu - \sigma)\\
&\qquad + (\sigma\boldsymbol{q}_\omega^\top - \boldsymbol{\pi}_{\boldsymbol{y}}^\top - \boldsymbol{\pi}^\top \boldsymbol{D}_\omega + \sigma\boldsymbol{x}^\top \boldsymbol{L}_\omega)\boldsymbol{y}.
\end{aligned}
$$

By weak duality, we have

$$
Q_\omega(\boldsymbol{x},\boldsymbol{\lambda},\sigma,\boldsymbol{\pi},\boldsymbol{\pi}_{\boldsymbol{y}},\pi_\gamma,\pi_\mu) := \min_{\boldsymbol{y},\mu,\gamma}\ L_\omega(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{y},\mu,\gamma,\sigma,\boldsymbol{\pi},\boldsymbol{\pi}_{\boldsymbol{y}},\pi_\gamma,\pi_\mu) \leq G_\omega(\boldsymbol{x},\boldsymbol{\lambda}),
$$

for all $\boldsymbol{\pi},\boldsymbol{\pi}_{\boldsymbol{y}} \geq \boldsymbol{0}$ and $\sigma,\pi_\gamma,\pi_\mu \geq 0$. Thus, problem (11) can be reformulated as:

$$
\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\theta}}\ & \boldsymbol{\lambda}^\top \boldsymbol{\vartheta}(\boldsymbol{x}) + \sum_{\omega\in[N]} \theta_\omega\\
\text{s.t.}\ \ & \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{\lambda} \geq \boldsymbol{0},\\
& \theta_\omega \geq Q_\omega(\boldsymbol{x},\boldsymbol{\lambda},\sigma,\boldsymbol{\pi},\boldsymbol{\pi}_{\boldsymbol{y}},\pi_\gamma,\pi_\mu), \quad \omega\in[N],\ \forall \boldsymbol{\pi},\boldsymbol{\pi}_{\boldsymbol{y}} \geq \boldsymbol{0},\ \sigma,\pi_\gamma,\pi_\mu \geq 0.
\end{aligned} \tag{30}
$$

Let us consider the Lagrangian dual problem

$$
\begin{aligned}
G_\omega(\boldsymbol{x},\boldsymbol{\lambda}) &= \max_{\boldsymbol{\pi},\boldsymbol{\pi}_{\boldsymbol{y}}\geq\boldsymbol{0},\sigma,\pi_\gamma,\pi_\mu\geq 0}\ Q_\omega(\boldsymbol{x},\boldsymbol{\lambda},\sigma,\boldsymbol{\pi},\boldsymbol{\pi}_{\boldsymbol{y}},\pi_\gamma,\pi_\mu)\\
&= \max_{\boldsymbol{\pi},\boldsymbol{\pi}_{\boldsymbol{y}}\geq\boldsymbol{0},\sigma,\pi_\gamma,\pi_\mu\geq 0}\ \Big\{\sigma\big(\boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\big) + \boldsymbol{\pi}^\top(\boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega) +\\
&\qquad \min_{\boldsymbol{y},\mu,\gamma}\ \gamma(\overline{p}_\omega - \pi_\gamma - \sigma) - \mu(\underline{p}_\omega + \pi_\mu - \sigma) + (\sigma\boldsymbol{q}_\omega^\top - \boldsymbol{\pi}_{\boldsymbol{y}}^\top - \boldsymbol{\pi}^\top \boldsymbol{D}_\omega + \sigma\boldsymbol{x}^\top \boldsymbol{L}_\omega)\boldsymbol{y}\Big\}\\
&= \max_{\boldsymbol{\pi},\boldsymbol{\pi}_{\boldsymbol{y}}\geq\boldsymbol{0},\sigma,\pi_\gamma,\pi_\mu\geq 0}\ \sigma\big(\boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\big) + \boldsymbol{\pi}^\top(\boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega)\\
&\qquad \text{s.t.}\ \ \overline{p}_\omega - \pi_\gamma - \sigma = 0,\\
&\qquad\qquad \underline{p}_\omega + \pi_\mu - \sigma = 0,\\
&\qquad\qquad \sigma\boldsymbol{q}_\omega^\top - \boldsymbol{\pi}_{\boldsymbol{y}}^\top - \boldsymbol{\pi}^\top \boldsymbol{D}_\omega + \sigma\boldsymbol{x}^\top \boldsymbol{L}_\omega = \boldsymbol{0}^\top.
\end{aligned} \tag{31}
$$

$$
\begin{aligned}
&= \max_{\boldsymbol{\pi}\geq \mathbf{0},\sigma\geq 0}\ \sigma\big(\boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\big) + \boldsymbol{\pi}^\top(\boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega) \\
&\qquad \text{s.t.}\quad \underline{p}_\omega \leq \sigma \leq \overline{p}_\omega, \\
&\qquad\qquad -\sigma \boldsymbol{q}_\omega^\top + \boldsymbol{\pi}^\top \boldsymbol{D}_\omega - \sigma \boldsymbol{x}^\top \boldsymbol{L}_\omega \leq \mathbf{0}^\top.
\end{aligned} \tag{32}
$$

Given $\hat{\boldsymbol{x}} \in \mathcal{X}$ and $\hat{\boldsymbol{\lambda}} \geq \mathbf{0}$, let $(\hat{\sigma}, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\pi}}_{\boldsymbol{y}}, \hat{\pi}_\gamma, \hat{\pi}_\mu)$ be a corresponding optimal Lagrange multipliers to Lagrangian dual problem (31) with zero duality gap. Hence, for $\omega \in [N]$, we have

$$
\begin{aligned}
&L_\omega(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{y}, \mu, \gamma, \hat{\sigma}, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\pi}}_{\boldsymbol{y}}, \hat{\pi}_\gamma, \hat{\pi}_\mu) \\
&= \hat{\sigma}\big(\boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\big) + \hat{\boldsymbol{\pi}}^\top(\boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega) + \gamma(\overline{p}_\omega - \hat{\pi}_\gamma - \hat{\sigma}) - \mu(\underline{p}_\omega + \hat{\pi}_\mu - \hat{\sigma}) \\
&\quad + (\hat{\sigma}\boldsymbol{q}_\omega^\top - \hat{\boldsymbol{\pi}}_{\boldsymbol{y}}^\top - \hat{\boldsymbol{\pi}}^\top \boldsymbol{D}_\omega + \hat{\sigma}\boldsymbol{x}^\top \boldsymbol{L}_\omega)\boldsymbol{y} \\
&= \hat{\sigma}\big(\boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\big) + \hat{\boldsymbol{\pi}}^\top(\boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega) + \hat{\sigma}(\boldsymbol{x} - \hat{\boldsymbol{x}})^\top \boldsymbol{L}_\omega \boldsymbol{y},
\end{aligned}
$$

where the second equality follows from the facts that $\overline{p}_\omega - \hat{\pi}_\gamma - \hat{\sigma} = 0$, $\underline{p}_\omega + \hat{\pi}_\mu - \hat{\sigma} = 0$, and $\hat{\sigma}\boldsymbol{q}_\omega^\top - \hat{\boldsymbol{\pi}}_{\boldsymbol{y}}^\top - \hat{\boldsymbol{\pi}}^\top \boldsymbol{D}_\omega + \hat{\sigma}\hat{\boldsymbol{x}}^\top \boldsymbol{L}_\omega = \mathbf{0}^\top$. Given that $L_\omega(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{y}, \mu, \gamma, \hat{\sigma}, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\pi}}_{\boldsymbol{y}}, \hat{\pi}_\gamma, \hat{\pi}_\mu)$ only depends on primal variables $\boldsymbol{y}$ and dual multipliers $(\hat{\sigma}, \hat{\boldsymbol{\pi}})$, we suppress (μ, γ) and $(\hat{\boldsymbol{\pi}}_{\boldsymbol{y}}, \hat{\pi}_\gamma, \hat{\pi}_\mu)$ from the arguments of L_ω. Moreover, L_ω depends on $\hat{\boldsymbol{x}}$ only through $(\hat{\sigma}, \hat{\boldsymbol{\pi}})$. Hence, we add $\hat{\boldsymbol{x}}$ to the arguments of L_ω, i.e., $L_\omega(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{y}, \mu, \gamma, \hat{\sigma}, \hat{\boldsymbol{\pi}}; \hat{\boldsymbol{x}})$, when needed for clarity. Similarly, we write $Q_\omega(\boldsymbol{x}, \boldsymbol{\lambda}, \hat{\sigma}, \hat{\boldsymbol{\pi}}; \hat{\boldsymbol{x}})$. Consequently, upon choosing $(\hat{\sigma}, \hat{\boldsymbol{\pi}})$, the Lagrangian dual function can be written as:

$$
\begin{aligned}
Q_\omega(\boldsymbol{x}, \boldsymbol{\lambda}, \hat{\sigma}, \hat{\boldsymbol{\pi}}; \hat{\boldsymbol{x}}) &= \min_{\boldsymbol{y}}\ L_\omega(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{y}, \hat{\sigma}, \hat{\boldsymbol{\pi}}; \hat{\boldsymbol{x}}) \\
&= \hat{\sigma}\big(\boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\big) + \hat{\boldsymbol{\pi}}^\top(\boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega) + \min_{\boldsymbol{y}}\ \hat{\sigma}(\boldsymbol{x} - \hat{\boldsymbol{x}})^\top \boldsymbol{L}_\omega \boldsymbol{y}.
\end{aligned} \tag{33}
$$

Given $\hat{\boldsymbol{x}} \in \mathcal{X}$ and $\hat{\boldsymbol{\lambda}} \geq \mathbf{0}$, and for the specific choice of $(\hat{\sigma}, \hat{\boldsymbol{\pi}})$, an optimal solution of the Lagrangian dual function only depends on variables $\boldsymbol{y}$, i.e., the variables that are bilinearly connected to $\boldsymbol{x}$ in the description of the objective function $\psi_0(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\xi}_\omega) = \boldsymbol{c}_\omega^\top \boldsymbol{x} + \boldsymbol{q}_\omega^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{L}_\omega \boldsymbol{y}$. Thus, given that $\hat{\sigma} \geq 0$, the condition

$$
(\boldsymbol{x} - \hat{\boldsymbol{x}})^\top \boldsymbol{L}_\omega(\boldsymbol{y} - \boldsymbol{y}_\omega^*) \geq 0, \quad \forall \boldsymbol{y} \tag{34}
$$

is sufficient and necessary for the optimality of $\boldsymbol{y}_\omega^*$ to the Lagrangian dual function (33). Given that $\hat{\sigma}(\boldsymbol{x} - \hat{\boldsymbol{x}})^\top \boldsymbol{L}_\omega \boldsymbol{y}$ is linear in $\boldsymbol{y}$, an optimal $\boldsymbol{y}_\omega^*$ is a boundary point of the space of $\boldsymbol{y}$. Suppose that Assumption **(A2)** holds and $\underline{\boldsymbol{y}}_\omega \leq \boldsymbol{y}_\omega \leq \overline{\boldsymbol{y}}_\omega$. Let $\boldsymbol{l}_{\omega,j}$ be the j-th column of matrix $\boldsymbol{L}_\omega$. Thus, if $(\boldsymbol{x} - \hat{\boldsymbol{x}})^\top \boldsymbol{l}_{\omega,j} \geq 0$, then $y_{\omega,j}^* = \underline{y}_{\omega,j}$. Otherwise, if $(\boldsymbol{x} - \hat{\boldsymbol{x}})^\top \boldsymbol{l}_{\omega,j} \leq 0$, then $y_{\omega,j}^* = \overline{y}_{\omega,j}$.

Let $\hat{\boldsymbol{y}}_\omega$ be a combination of such lower and upper bounds on variables $\boldsymbol{y}_\omega$ that attains Lagrangian dual function (33). Thus,

$$
\theta_\omega \geq \hat{\sigma}\big(\boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\big) + \hat{\boldsymbol{\pi}}^\top(\boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega) + \hat{\sigma}(\boldsymbol{x} - \hat{\boldsymbol{x}})^\top \boldsymbol{L}_\omega \hat{\boldsymbol{y}}_\omega
$$

is a valid linear inequality for (30) on

$$
\mathcal{S}_\omega := \left\{ \boldsymbol{x} \;\middle|\; \begin{array}{l} (\boldsymbol{x} - \hat{\boldsymbol{x}})^\top \boldsymbol{l}_{\omega,j} \geq 0,\ j \in J_\omega^+, \\ (\boldsymbol{x} - \hat{\boldsymbol{x}})^\top \boldsymbol{l}_{\omega,j} \leq 0,\ j \in J_\omega^- \end{array} \right\}, \tag{35}
$$

where $J_\omega^+ := \{j : \hat{y}_{\omega,j} = \underline{y}_{\omega,j}\}$ and $J_\omega^- := \{j : \hat{y}_{\omega,j} = \overline{y}_{\omega,j}\}$. Let Q be a finite index set, enumerating all combinations of such lower and upper bounds on variables $\boldsymbol{y}_\omega$. Appending the subscript $q \in Q$ to $\hat{\boldsymbol{y}}$, J_ω^+, and J_ω^-, we define

$$
\mathcal{K}_{\omega,q} := \left\{ (\boldsymbol{x}, \boldsymbol{\lambda}, \theta_\omega) \;\middle|\; \begin{array}{l} \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{d},\ \boldsymbol{x} \geq \mathbf{0},\ \boldsymbol{\lambda} \geq \mathbf{0}, \\ \theta_\omega \geq \hat{\sigma}\big(\boldsymbol{c}_\omega^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)\big) + \\ \qquad \hat{\boldsymbol{\pi}}^\top(\boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega) + \hat{\sigma}(\boldsymbol{x} - \hat{\boldsymbol{x}})^\top \boldsymbol{L}_\omega \hat{\boldsymbol{y}}_{\omega,q}, \\ (\boldsymbol{x} - \hat{\boldsymbol{x}})^\top \boldsymbol{l}_{\omega,j} \geq 0,\ j \in J_{\omega,q}^+, \\ (\boldsymbol{x} - \hat{\boldsymbol{x}})^\top \boldsymbol{l}_{\omega,j} \leq 0,\ j \in J_{\omega,q}^- \end{array} \right\}. \tag{36}
$$

Given that Lagrangian dual function (33) attains its optimal solution at $\hat{\boldsymbol{y}}_{\omega,q}$ for some $q \in Q$, the feasible region of (30) is a subset of the following disjunctive set:

$$\mathcal{T}_\omega := \left\{ (\boldsymbol{x}, \boldsymbol{\lambda}, \theta_\omega) : \bigcup_{q \in Q} \mathcal{K}_{\omega,q} \right\}. \tag{37}$$

Following Balas [2, Theorem 3.1], a valid inequality for the convex hull of $\mathcal{T}_\omega$, $\operatorname{conv}(\mathcal{T}_\omega)$, may be represented in the form

$$\boldsymbol{\alpha}_{\boldsymbol{x},\omega}^\top \boldsymbol{x} + \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}^\top \boldsymbol{\lambda} + \alpha_{\theta,\omega}\theta_\omega \geq \rho_\omega, \tag{38}$$

where $(\boldsymbol{\alpha}_{\boldsymbol{x},\omega}, \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}, \alpha_{\theta,\omega}, \rho_\omega)$ is an element of the following polyhedral set:

$$\mathcal{W}_\omega := \left\{ (\boldsymbol{\alpha}_{\boldsymbol{x},\omega}, \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}, \alpha_{\theta,\omega}, \rho_\omega) \;\middle|\; \begin{array}{l} \exists\, \{\delta^+_{\omega,j,q} \geq 0 : j \in J^+_{\omega,q}\}, \\ \quad \{\delta^-_{\omega,j,q} \geq 0 : j \in J^-_{\omega,q}\}, \\ \quad \delta_{\theta,q} \geq 0,\ \boldsymbol{\delta}_{\boldsymbol{x},q} \geq \boldsymbol{0},\ q \in Q,\ \text{s.t.} \\ \boldsymbol{A}^\top \boldsymbol{\delta}_{\boldsymbol{x},q} - \delta_{\theta,q}\hat{\sigma}(\boldsymbol{c} + \boldsymbol{L}_\omega \hat{\boldsymbol{y}}_{\omega,q}) - \delta_{\theta,q}\boldsymbol{B}_\omega^\top \hat{\boldsymbol{\pi}} + \\ \quad \sum_{j \in J^+_{\omega,q}} \delta^+_{\omega,j,q}\boldsymbol{l}_{\omega,j} - \\ \quad \sum_{j \in J^-_{\omega,q}} \delta^-_{\omega,j,q}\boldsymbol{l}_{\omega,j} \leq \boldsymbol{\alpha}_{\boldsymbol{x},\omega},\ q \in Q, \\ \delta_{\theta,q}\hat{\sigma}\boldsymbol{g}(\boldsymbol{\xi}_\omega) \leq \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega},\ q \in Q, \\ \delta_{\theta,q} \leq \alpha_{\theta,\omega},\ q \in Q, \\ \boldsymbol{\delta}_{\boldsymbol{x},q}^\top \boldsymbol{d} + \delta_{\theta,q}\hat{\boldsymbol{\pi}}^\top \boldsymbol{b}_\omega - \delta_{\theta,q}\hat{\sigma}\hat{\boldsymbol{x}}^\top \boldsymbol{L}_\omega \hat{\boldsymbol{y}}_{\omega,q} + \\ \quad \sum_{j \in J^+_{\omega,q}} \delta^+_{\omega,j,q}\hat{\boldsymbol{x}}^\top \boldsymbol{l}_{\omega,j} - \\ \quad \sum_{j \in J^-_{\omega,q}} \delta^-_{\omega,j,q}\hat{\boldsymbol{x}}^\top \boldsymbol{l}_{\omega,j} \geq \rho_\omega,\ q \in Q, \end{array} \right\}, \tag{39}$$

The polyhedral set $\mathcal{W}_\omega$ is the reverse polar cone of $\operatorname{conv}(\mathcal{T}_\omega)$, i.e., the cone characterizing all valid inequalities for $\operatorname{conv}(\mathcal{T}_\omega)$. Moreover, (38) is a facet of $\operatorname{conv}(\mathcal{T}_\omega)$ if and only if $(\boldsymbol{\alpha}_{\boldsymbol{x},\omega}, \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}, \alpha_{\theta,\omega}, \rho_\omega)$ is an extreme ray of $\mathcal{W}_\omega$. Such a facet can be identified by solving the following cut-generation linear program (CGLP):

$$\min_{(\boldsymbol{\alpha}_{\boldsymbol{x},\omega}, \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}, \alpha_{\theta,\omega}, \rho_\omega) \in \mathcal{W}_\omega} \ \boldsymbol{\alpha}_{\boldsymbol{x},\omega}^\top \hat{\boldsymbol{x}} + \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}^\top \hat{\boldsymbol{\lambda}} + \alpha_{\theta,\omega}^\top \hat{\theta}_\omega - \rho_\omega. \tag{40}$$

If the optimal value of (40) is nonnegative, then $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}, \hat{\theta}_\omega) \in \operatorname{conv}(\mathcal{T}_\omega)$. Otherwise, if the optimal value of (40) is negative and $(\boldsymbol{\alpha}_{\boldsymbol{x},\omega}, \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}, \alpha_{\theta,\omega}, \rho_\omega)$ is an optimal solution, then a disjunctive inequality in the form of (38) is valid for $\operatorname{conv}(\mathcal{T}_\omega)$, which cuts off $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}, \hat{\theta}_\omega)$.

As $\mathcal{W}_\omega$ is a cone, a normalizing constraint like $\sum_{q \in Q} \delta_{\theta,q} = 1$, or $\sum_{q \in Q} \Big(\sum_{j \in J^+_{\omega,q}} \delta^+_{\omega,j,q} + \sum_{j \in J^-_{\omega,q}} \delta^-_{\omega,j,q} + \delta_{\theta,q} + \boldsymbol{\delta}_{\boldsymbol{x},q}^\top \boldsymbol{e} \Big) = 1$ can be added to $\mathcal{W}_\omega$ once solving the CGLP (40). Note that if the normalizing constraint $\sum_{q \in Q} \delta_{\theta,q} = 1$ is used, then the CGLP allows for those facets of $\operatorname{conv}(\mathcal{T}_\omega)$ that have a positive coefficient $\alpha_{\theta,\omega}$ for θ_ω, as $\alpha_{\theta,\omega} \geq \max_{q \in Q} \delta_{\theta,q} > 0$ by the constraints in $\mathcal{W}_\omega$. Hence, a valid inequality is given by

$$\theta_\omega + \left[\frac{\boldsymbol{\alpha}_{\boldsymbol{x},\omega}}{\alpha_{\theta,\omega}}\right]^\top \boldsymbol{x} + \left[\frac{\boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}}{\alpha_{\theta,\omega}}\right]^\top \boldsymbol{\lambda} \geq \frac{\rho_\omega}{\alpha_{\theta,\omega}}.$$

Putting these all together, one can form a restricted master problem for (1) after t iterations as follows:

$$z^t = \min_{(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\theta}) \in \mathcal{S}^t} \ \boldsymbol{\lambda}^\top \boldsymbol{\vartheta}(\boldsymbol{x}) + \sum_{\omega \in [N]} \theta_\omega, \tag{41}$$

Algorithm 4: $\texttt{SepCuts}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}, \hat{\theta}_\omega; \hat{\boldsymbol{\pi}}, \hat{\sigma})$.

Input: $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}, \hat{\theta}_\omega)$ and $(\hat{\boldsymbol{\pi}}, \hat{\sigma})$.

Output: $(\texttt{viol}, \boldsymbol{\alpha}_{\boldsymbol{x},\omega}, \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}, \alpha_{\theta,\omega}, \rho_\omega)$. If a valid inequality $\boldsymbol{\alpha}_{\boldsymbol{x},\omega}^\top \boldsymbol{x} + \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}^\top \boldsymbol{\lambda} + \alpha_{\theta,\omega}\theta_\omega \geq \rho_\omega$ is found that is violated by $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}}, \hat{\theta}_\omega)$, then return viol=TRUE, $\boldsymbol{\alpha}_{\boldsymbol{x},\omega}$, $\boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}$, $\alpha_{\theta,\omega}$, and ρ_ω. Otherwise, return viol=FALSE, $\boldsymbol{\alpha}_{\boldsymbol{x},\omega} = \mathbf{0}$, $\boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega} = \mathbf{0}$, $\alpha_{\theta,\omega} = 0$, and $\rho_\omega = 0$.

Initialization: viol $\leftarrow$ FALSE, $\boldsymbol{\alpha}_{\boldsymbol{x},\omega} = \mathbf{0}$, $\boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega} = \mathbf{0}$, $\alpha_{\theta,\omega} = 0$, and $\rho_\omega = 0$.
Given $(\hat{\boldsymbol{\pi}}, \hat{\sigma})$, form $\mathcal{W}_\omega$ as in (39), with a proper a normalization constraint.
Let $(\boldsymbol{\alpha}^*_{\boldsymbol{x},\omega}, \boldsymbol{\alpha}^*_{\boldsymbol{\lambda},\omega}, \alpha^*_{\theta,\omega}, \rho^*_\omega)$ be an optimal solution to the CGLP

$$\hat{z} = \min_{(\boldsymbol{\alpha}_{\boldsymbol{x},\omega}, \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}, \alpha_{\theta,\omega}, \rho_\omega) \in \mathcal{W}_\omega} \ \boldsymbol{\alpha}_{\boldsymbol{x},\omega}^\top \hat{\boldsymbol{x}} + \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}^\top \hat{\boldsymbol{\lambda}} + \alpha_{\theta,\omega}^\top \hat{\theta}_\omega - \rho_\omega.$$

if $\hat{z} < 0$ **then**
  viol $\leftarrow$ TRUE and let $\boldsymbol{\alpha}_{\boldsymbol{x},\omega} = \boldsymbol{\alpha}^*_{\boldsymbol{x},\omega}$, $\boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega} = \boldsymbol{\alpha}^*_{\boldsymbol{\lambda},\omega}$, $\alpha_{\theta,\omega} = \alpha^*_{\theta,\omega}$, and $\rho_\omega = \rho^*_\omega$.
end

where

$$\mathcal{S}^t = \left\{ (\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\theta}) \,\middle|\, \begin{array}{l} \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{d}, \ \boldsymbol{x} \geq \mathbf{0}, \ \boldsymbol{\lambda} \geq \mathbf{0}, \\ \boldsymbol{x}^\top \boldsymbol{\alpha}^k_{\boldsymbol{x},\omega} + \boldsymbol{\lambda}^\top \boldsymbol{\alpha}^k_{\boldsymbol{\lambda},\omega} + \\ \qquad \theta_\omega \alpha^k_{\theta,\omega} \geq \rho^k_\omega, \ \omega \in [N], \ k \in \{0\} \cup [t-1] \end{array} \right\}. \tag{42}$$

Here, $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t, \boldsymbol{\theta}^t)$ is a solution obtained from solving (41) at iteration t, and $(\boldsymbol{\alpha}^k_{\boldsymbol{x},\omega}, \boldsymbol{\alpha}^k_{\boldsymbol{\lambda},\omega}, \alpha^k_{\theta,\omega}, \rho^k_\omega)$ are the corresponding cut coefficients, $\omega \in [N]$ and $k \in \{0\} \cup [t-1]$. A general framework of a finitely-convergent disjunctive cutting plane scheme to solve (1) with a nonconvex recourse function and a generic decision-dependent ambiguity set is given in Algorithm 5.

Theorem 5. *Suppose that Assumptions* **(A1)**–**(A3)** *and* **(A5)** *hold. For $\omega \in [N]$, suppose that $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ is a nonconvex recourse function defined in* (2)*. Then, Algorithm 5 generates an ϵ-optimal solution to* (1) *in a finite number of iterations.*

To prove Theorem 5, we present some lemmas. The next two lemmas hold for problem (1) with a generic recourse function (2).

Lemma 2. *Suppose that Assumptions* **(A1)**-**(A3)** *hold. For a fixed $\boldsymbol{x} \in \mathcal{X}$ and $\omega \in [N]$, $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$, defined in* (9)*, is a proper, convex, continuous function in $\boldsymbol{\lambda}$ on $\mathbb{R}^s$.*

Proof. Note that by the boundedness of $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ (implied by Assumption **A2**), we have $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}) > -\infty$ for all $\boldsymbol{\lambda} \in \mathbb{R}^s$ and there exists $\boldsymbol{\lambda} \in \mathbb{R}^s$ with $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda}) < \infty$, e.g., $\boldsymbol{\lambda} = \mathbf{0}$; proving $G_\omega(\boldsymbol{x}, \cdot)$ is a proper function. In addition, by the convexity of $\varphi_\omega[\cdot]$ from Lemma 1 and linearity of $h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{\xi}_\omega)$ in $\boldsymbol{\lambda}$, $G_\omega(\boldsymbol{x}, \cdot)$ is a convex function, and hence, continuous on $\mathbb{R}^s$. □

Let $\mathcal{L}$ denote the set of optimal multipliers $\boldsymbol{\lambda}$ in problem (8). The next lemma establishes the boundedness of $\mathcal{L}$.

Lemma 3. *Suppose that Assumptions* **(A1)**–**(A3)** *hold. Then, the set of optimal multipliers $\boldsymbol{\lambda}$ in problem* (8)*, $\mathcal{L}$, is compact.*

Proof. For a fixed $\boldsymbol{x} \in \mathcal{X}$, following the proof of Theorem 1, problem $\max_{\boldsymbol{p} \in \mathcal{P}(\boldsymbol{x})} \mathbb{E}_{\boldsymbol{p}} [h(\boldsymbol{x}, \boldsymbol{\xi})]$ can be reformulated as the minimization problem $\min_{\boldsymbol{\lambda} \geq \mathbf{0}} \ \boldsymbol{\lambda}^\top \boldsymbol{\vartheta}(\boldsymbol{x}) + \sum_{\omega \in [N]} G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$. Given that for $\boldsymbol{x} \in \mathcal{X}$, $\sum_{\omega \in [N]} G_\omega(\boldsymbol{x}, \cdot)$ is a proper, convex, continuous function by Lemma 2, the minimum of

Algorithm 5: Cutting-plane algorithm for problem (1) with a nonconvex recourse function and a generic decision-dependent ambiguity set.

Input: An initial solution $(\boldsymbol{x}^0, \boldsymbol{\lambda}^0, \boldsymbol{\theta}^0)$ and $\epsilon > 0$ for the optimality tolerance.
Output: An ϵ-optimal solution and the ϵ-optimal value.

Initialization: Set $t \leftarrow 1$, $\Omega \leftarrow \emptyset$, $\text{LB} \leftarrow -\infty$, $\text{UB} \leftarrow +\infty$. Add initial cuts (e.g., $\theta_\omega \geq 0$), if available, to $\mathcal{S}^0$.
while $\text{UB} - \frac{\epsilon}{2} > \text{LB}$ **do**
  for ***each*** $\omega \in [N]$ **do**
    Obtain $G_\omega(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1})$ by solving (29) and $(\boldsymbol{\pi}_\omega^{t-1}, \sigma_\omega^{t-1})$ by solving (32) at $(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1})$.
    Call the procedure $\texttt{SepCuts}(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1}, \theta_\omega^{t-1}; \boldsymbol{\pi}_\omega^{t-1}, \sigma_\omega^{t-1})$ to obtain $(\texttt{viol}, \boldsymbol{\alpha}_{\boldsymbol{x},\omega}, \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}, \alpha_{\theta,\omega}, \rho_\omega)$.
    if $\texttt{viol}$=*TRUE* **then**
      $\Omega \leftarrow \Omega \cup \{\omega\}$.
    end
  end
  $\text{UB} \leftarrow \min\{\text{UB}, \boldsymbol{\vartheta}(\boldsymbol{x}^{t-1})^\top \boldsymbol{\lambda}^{t-1} + \sum_{\omega \in [N]} G_\omega(\boldsymbol{x}^{t-1}, \boldsymbol{\lambda}^{t-1})\}$.
  Let $\mathcal{S}^t \leftarrow \mathcal{S}^{t-1} \cap \left\{(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\theta}) \,\middle|\, \boldsymbol{\alpha}_{\boldsymbol{x},\omega}^\top \boldsymbol{x} + \boldsymbol{\alpha}_{\boldsymbol{\lambda},\omega}^\top \boldsymbol{\lambda} + \alpha_{\theta,\omega}\theta_\omega \geq \rho_\omega,\ \omega \in \Omega\right\}$.
  Solve restricted master problem (41) using $\mathcal{S}^t$ and obtain an $\frac{\epsilon}{2}$-optimal solution $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t, \boldsymbol{\theta}^t)$.
  Let $\text{LB} \leftarrow \boldsymbol{\vartheta}(\boldsymbol{x}^t)^\top \boldsymbol{\lambda}^t + \sum_{\omega \in [N]} \theta_\omega^t$.
  Set $t \leftarrow t + 1$, $\Omega \leftarrow \emptyset$.
end
return $(\boldsymbol{x}^t, \boldsymbol{\lambda}^t)$ and UB.

this convex minimization problem is attained on a finite $\mathbf{0} \leq \boldsymbol{\lambda}$ for all $\boldsymbol{x} \in \mathcal{X}$. This completes the proof. □

For $\omega \in [N]$ and a fixed $(\boldsymbol{x}, \boldsymbol{\lambda}) \in \mathcal{X} \times \mathcal{L}$, let $\mathcal{F}_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ and $\mathcal{E}_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ denote the (primal) feasible region of problem (29) and the (dual) feasible region of problem (32), respectively. Moreover, let $\mathcal{D}_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ denote the set of optimal solutions for problem (32).

Lemma 4. *Suppose that Assumptions* **(A1)-(A3)** *hold. For $\omega \in [N]$, we have:*

i. $\mathcal{D}_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ is compact for a fixed $(\boldsymbol{x}, \boldsymbol{\lambda}) \in \mathcal{X} \times \mathcal{L}$,

ii. $\mathcal{F}_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ is compact for a fixed $(\boldsymbol{x}, \boldsymbol{\lambda}) \in \mathcal{X} \times \mathcal{L}$,

iii. The set-valued function $\mathcal{E}_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ is continuous in $(\boldsymbol{x}, \boldsymbol{\lambda})$ on $\mathcal{X} \times \mathcal{L}$,

iv. The real-valued function $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$, as defined in (29), is l.s.c. in $(\boldsymbol{x}, \boldsymbol{\lambda})$ on $\mathcal{X} \times \mathcal{L}$, and

v. The set-valued function $\mathcal{D}_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ is u.s.c. in $(\boldsymbol{x}, \boldsymbol{\lambda})$ on $\mathcal{X} \times \mathcal{L}$.

Proof. Consider a fixed $(\boldsymbol{x}, \boldsymbol{\lambda}) \in \mathcal{X} \times \mathcal{L}$. Recall that $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$, defined in (9), is well defined. Thus, $\mathcal{F}_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ is nonempty. Consequently, by a similar argument as that in Remark 1, $\mathcal{D}_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ is a bounded (and closed) set. Now, note that for a fixed $(\boldsymbol{x}, \boldsymbol{\lambda}) \in \mathcal{X} \times \mathcal{L}$, $\mathcal{F}_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ is closed and bounded given that $\mathcal{Y}(\boldsymbol{x}, \boldsymbol{\xi}_\omega)$ is bounded by Assumption **(A2)**, and the fact that μ and γ are bounded by

$|h(\boldsymbol{x},\boldsymbol{\xi}_\omega)-\boldsymbol{\lambda}^\top\boldsymbol{g}(\boldsymbol{\xi}_\omega)|$ from above. By the boundedness of $\mathcal{F}_\omega(\boldsymbol{x},\boldsymbol{\lambda})$ and a direct application of [72, Corollary 11], $\mathcal{E}_\omega(\boldsymbol{x},\boldsymbol{\lambda})$ is continuous in $(\boldsymbol{x},\boldsymbol{\lambda})$ on $\mathcal{X}\times\mathcal{L}$. By the continuity of $\mathcal{E}_\omega(\boldsymbol{x},\boldsymbol{\lambda})$ and the direct application of [72, Theorem 2], we have that $G_\omega(\boldsymbol{x},\boldsymbol{\lambda})$ is l.s.c. in $(\boldsymbol{x},\boldsymbol{\lambda})$ on $\mathcal{X}\times\mathcal{L}$. Moreover, $\mathcal{D}_\omega(\boldsymbol{x},\boldsymbol{\lambda})$ is equivalent to

$$\Big\{(\boldsymbol{\pi},\sigma)\in\mathcal{E}_\omega(\boldsymbol{x},\boldsymbol{\lambda})\,\Big|\,G_\omega(\boldsymbol{x},\boldsymbol{\lambda})=\sigma\big(\boldsymbol{c}_\omega^\top\boldsymbol{x}-\boldsymbol{\lambda}^\top\boldsymbol{g}(\boldsymbol{\xi}_\omega)\big)+\boldsymbol{\pi}^\top(\boldsymbol{B}_\omega\boldsymbol{x}+\boldsymbol{b}_\omega)\Big\}.$$

Since the objective function of (32), $\sigma\big(\boldsymbol{c}_\omega^\top\boldsymbol{x}-\boldsymbol{\lambda}^\top\boldsymbol{g}(\boldsymbol{\xi}_\omega)\big)+\boldsymbol{\pi}^\top(\boldsymbol{B}_\omega\boldsymbol{x}+\boldsymbol{b}_\omega)$, is continuous on $\mathcal{E}_\omega(\boldsymbol{x},\boldsymbol{\lambda})\times\mathcal{X}\times\mathcal{L}$ and the set-valued function $\mathcal{E}_\omega(\boldsymbol{x},\boldsymbol{\lambda})$ is continuous on $\mathcal{X}\times\mathcal{L}$, the direct application of [39, Theorem 1.5] shows the set-valued function $\mathcal{D}_\omega(\boldsymbol{x},\boldsymbol{\lambda})$ is u.s.c. in $(\boldsymbol{x},\boldsymbol{\lambda})$ on $\mathcal{X}\times\mathcal{L}$. □

of Theorem 5. To prove the finite convergence of Algorithm 5, we need to show that the "while" loop terminates in a finite number of iterations, generating an ϵ-optimal solution to (1).

By contradiction, suppose that the "while" loop does not terminate in a finite number of iterations. Let $\{(\boldsymbol{x}^t,\boldsymbol{\lambda}^t,\boldsymbol{\theta}^t)\}$ be the sequence of iterates generated in Line 12 and $\{(\boldsymbol{\pi}^t_\omega,\sigma^t_\omega)\in\mathcal{D}_\omega(\boldsymbol{x}^t,\boldsymbol{\lambda}^t):\omega\in[N]\}$ be the sequence of dual multipliers generated in Line 4. We show that $\theta^t_\omega\geq G_\omega(\boldsymbol{x}^t,\boldsymbol{\lambda}^t)-\frac{\epsilon}{2}$, $\omega\in[N]$, for all sufficiently large t. This implies that we have

$$\mathrm{LB}=\boldsymbol{\lambda}^t\boldsymbol{\vartheta}(\boldsymbol{x}^t)+\sum_{\omega\in[N]}\theta^t_\omega\geq\boldsymbol{\lambda}^t\boldsymbol{\vartheta}(\boldsymbol{x}^t)+\sum_{\omega\in[N]}G_\omega(\boldsymbol{x}^t,\lambda^t)-\frac{\epsilon}{2}=\mathrm{UB}-\frac{\epsilon}{2},$$

contradicting that the "while" loop does not terminate in a finite number of iterations. Given Assumption **(A5)**, we have $\mathrm{LB}\leq z^t+\frac{\epsilon}{2}$. Thus, Algorithm 5 returns the ϵ-optimal value UB because $\mathrm{UB}\leq z^t+\epsilon$, with the corresponding ϵ-optimal solution $(\boldsymbol{x}^t,\boldsymbol{\lambda}^t)$.

Now, we prove that $\theta^t_\omega\geq G_\omega(\boldsymbol{x}^t,\boldsymbol{\lambda}^t)-\frac{\epsilon}{2}$, $\omega\in[N]$, for all sufficiently large t. Note that $\{\boldsymbol{x}^t\}$ and $\{\boldsymbol{\lambda}^t\}$ are bounded by the compactness of $\mathcal{X}$ and $\mathcal{L}$ (by Lemma 3), respectively. Moreover, $\{\boldsymbol{\theta}^t\}$ is a nondecreasing bounded sequence from below given that θ^t_ω is an underestimator of $G_\omega(\boldsymbol{x},\boldsymbol{\lambda})$, $\omega\in[N]$, and accumulation of constraints in $\mathcal{S}^t$. In addition, $\{\boldsymbol{\theta}^t\}$ is bounded from above given that $\theta^t_\omega\leq G_\omega(\boldsymbol{x}^t,\boldsymbol{\lambda}^t)$, $\omega\in[N]$. Consequently, there is a convergent subsequence, say $\mathcal{K}$, $\{(\boldsymbol{x}^t,\boldsymbol{\lambda}^t,\boldsymbol{\theta}^t)\}_{t\in\mathcal{K}}$. In addition, the associated sequence of optimal dual multipliers $\{(\boldsymbol{\pi}^t_\omega,\sigma^t_\omega):\omega\in[N]\}_{t\in\mathcal{K}}$ is bounded by Lemma 4.i.; hence, there is a convergent subsequence on $\mathcal{K}'\subseteq\mathcal{K}$. Let $\{(\overline{\boldsymbol{x}},\overline{\boldsymbol{\lambda}},\overline{\boldsymbol{\theta}})\}\in\mathcal{X}\times\mathcal{L}$ (by closedness) and $\{(\overline{\boldsymbol{\pi}}_\omega,\overline{\sigma}_\omega):\omega\in[N]\}$ be limit point of $\{(\boldsymbol{x}^t,\boldsymbol{\lambda}^t,\boldsymbol{\theta}^t)\}$ and $\{(\boldsymbol{\pi}^t_\omega,\sigma^t_\omega):\omega\in[N]\}$ on $\mathcal{K}'$, respectively. As $\mathcal{D}_\omega(\boldsymbol{x},\boldsymbol{\lambda})$, $\omega\in[N]$, is u.s.c. at $(\overline{\boldsymbol{x}},\overline{\boldsymbol{\lambda}})$ by Lemma 4.v., we have that $(\overline{\boldsymbol{\pi}}_\omega,\overline{\sigma}_\omega)\in\mathcal{D}_\omega(\overline{\boldsymbol{x}},\overline{\boldsymbol{\lambda}})$, $\omega\in[N]$. Let $\{(\boldsymbol{y}^t_\omega,\mu^t_\omega,\gamma^t_\omega)\in\mathcal{F}_\omega(\boldsymbol{x}^t,\boldsymbol{\lambda}^t):\omega\in[N]\}$, be the sequence of associated optimal primal solutions to problem (29). Given that $\mathcal{F}_\omega(\boldsymbol{x}^t,\boldsymbol{\lambda}^t)$ is compact by Lemma 4.ii., there is a convergent subsequence on $\mathcal{K}''\subseteq\mathcal{K}'$ with a limit point $\{(\overline{\boldsymbol{y}}_\omega,\overline{\mu}_\omega,\overline{\gamma}_\omega):\omega\in[N]\}$ on $\mathcal{K}''$.

We now claim that $L_\omega(\overline{\boldsymbol{x}},\overline{\boldsymbol{\lambda}},\boldsymbol{y}_{\omega,q},\overline{\sigma}_\omega,\overline{\boldsymbol{\pi}}_\omega;\overline{\boldsymbol{x}})=L_\omega(\overline{\boldsymbol{x}},\overline{\boldsymbol{\lambda}},\overline{\boldsymbol{y}}_\omega,\overline{\sigma}_\omega,\overline{\boldsymbol{\pi}}_\omega;\overline{\boldsymbol{x}})$ for every $q\in Q$. Note that for every $(\boldsymbol{x},\boldsymbol{\lambda})\in\mathcal{X}\times\mathcal{L}$, we have $\lim_{t\to\infty}L_\omega(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{y}_{\omega,q},\sigma^t_\omega,\boldsymbol{\pi}^t_\omega;\boldsymbol{x}^t)=\overline{\sigma}_\omega\big(\boldsymbol{c}_\omega^\top\boldsymbol{x}-\boldsymbol{\lambda}^\top\boldsymbol{g}(\boldsymbol{\xi}_\omega)\big)+\overline{\boldsymbol{\pi}}_\omega^\top(\boldsymbol{B}_\omega\boldsymbol{x}+\boldsymbol{b}_\omega)+\overline{\sigma}_\omega(\boldsymbol{x}-\overline{\boldsymbol{x}})^\top\boldsymbol{L}_\omega\boldsymbol{y}_{\omega,q}=L_\omega(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{y}_{\omega,q},\overline{\sigma}_\omega,\overline{\boldsymbol{\pi}}_\omega;\overline{\boldsymbol{x}})$ by the continuity of $L_\omega(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{y}_{\omega,q},\hat{\sigma},\hat{\boldsymbol{\pi}};\hat{\boldsymbol{x}})$ at $(\hat{\sigma},\hat{\boldsymbol{\pi}};\hat{\boldsymbol{x}})$. Thus, $L_\omega(\overline{\boldsymbol{x}},\overline{\boldsymbol{\lambda}},\boldsymbol{y}_{\omega,q},\overline{\sigma}_\omega,\overline{\boldsymbol{\pi}}_\omega;\overline{\boldsymbol{x}})=\overline{\sigma}_\omega\big(\boldsymbol{c}_\omega^\top\overline{\boldsymbol{x}}-\overline{\boldsymbol{\lambda}}^\top\boldsymbol{g}(\boldsymbol{\xi}_\omega)\big)+\overline{\boldsymbol{\pi}}_\omega^\top(\boldsymbol{B}_\omega\overline{\boldsymbol{x}}+\boldsymbol{b}_\omega)=L_\omega(\overline{\boldsymbol{x}},\overline{\boldsymbol{\lambda}},\overline{\boldsymbol{y}}_\omega,\overline{\sigma}_\omega,\overline{\boldsymbol{\pi}}_\omega;\overline{\boldsymbol{x}})$ for every $q\in Q$.

Given the validity of optimality cut generated at Line 5 for the corresponding set conv (T_ω), we have $\theta_\omega\geq\sigma^t_\omega\big(\boldsymbol{c}_\omega^\top\boldsymbol{x}-\boldsymbol{\lambda}^\top\boldsymbol{g}(\boldsymbol{\xi}_\omega)\big)+(\boldsymbol{B}_\omega\boldsymbol{x}+\boldsymbol{b}_\omega)^\top\boldsymbol{\pi}^t_\omega+\sigma^t_\omega(\boldsymbol{x}-\boldsymbol{x}^t)^\top\boldsymbol{L}_\omega\boldsymbol{y}_{\omega,q}$ for some $q\in Q$. Moreover, given the accumulation of cuts, we have $\theta^{t+1}_\omega\geq\sigma^t_\omega\big(\boldsymbol{c}_\omega^\top\boldsymbol{x}^{t+1}-\boldsymbol{g}(\boldsymbol{\xi}_\omega)^\top\boldsymbol{\lambda}^{t+1}\big)+(\boldsymbol{B}_\omega\boldsymbol{x}^{t+1}+\boldsymbol{b}_\omega)^\top\boldsymbol{\pi}^t_\omega+\sigma^t_\omega(\boldsymbol{x}^{t+1}-\boldsymbol{x}^t)^\top\boldsymbol{L}_\omega\boldsymbol{y}_{\omega,q}$. Thus, taking the limit on $\mathcal{K}''$, we have $\overline{\theta}_\omega\geq L_\omega(\overline{\boldsymbol{x}},\overline{\boldsymbol{\lambda}},\boldsymbol{y}_{\omega,q},\overline{\sigma}_\omega,\overline{\boldsymbol{\pi}}_\omega;\overline{\boldsymbol{x}})$. And, using the above claim yields $\overline{\theta}_\omega\geq L_\omega(\overline{\boldsymbol{x}},\overline{\boldsymbol{\lambda}},\overline{\boldsymbol{y}}_\omega,\overline{\sigma}_\omega,\overline{\boldsymbol{\pi}}_\omega;\overline{\boldsymbol{x}})$, $\omega\in[N]$. Now, given that $\{(\overline{\boldsymbol{\pi}}_\omega,\overline{\sigma}_\omega)\in\mathcal{D}_\omega(\overline{\boldsymbol{x}},\overline{\boldsymbol{\lambda}}):\omega\in[N]\}$, by strong duality we have $G_\omega(\overline{\boldsymbol{x}},\overline{\boldsymbol{\lambda}})=L_\omega(\overline{\boldsymbol{x}},\overline{\boldsymbol{\lambda}},\overline{\boldsymbol{y}}_\omega,\overline{\sigma}_\omega,\overline{\boldsymbol{\pi}}_\omega;\overline{\boldsymbol{x}})$. Hence, $\overline{\theta}_\omega\geq$

$G_\omega(\overline{\boldsymbol{x}}, \overline{\boldsymbol{\lambda}})$. Finally, because $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$ is l.s.c. by Lemma 4.iv., we have $\theta^t_\omega \geq G_\omega(\boldsymbol{x}^t, \boldsymbol{\lambda}^t) - \frac{\epsilon}{2}$, $\omega \in [N]$, for all sufficiently large t. This completes the proof. □

Recall that when $\mathcal{P}(\boldsymbol{x})$ is decision-independent, i.e., $\vartheta(\boldsymbol{x}) = \boldsymbol{\vartheta}$, problem (11) reduces to a two-stage stochastic mixed-integer program with (convex or nonconvex) recourse. Thus, a modified version of Algorithm 5, by just changing $\vartheta(\boldsymbol{x})$ to $\boldsymbol{\vartheta}$, yields a decomposition-based cutting plane algorithm to solve the resulting DRO problem with a decision-independent ambiguity set to ϵ-optimality. We end this section with a remark on how Algorithm 5 is of independent interest to solve two-stage stochastic programs with a random decision-dependent recourse matrix (i.e., bilinear stochasticity on the left-hand side) or a bilinear objective function.

Remark 6. *Following similar ideas as those presented to develop Algorithm 5, a modified algorithm may be developed to find an ϵ-optimal solution to a two-stage stochastic mixed-integer linear program with a nonconvex recourse as follows:*

$$\min_{\boldsymbol{x} \in \mathcal{X}} \; \mathbb{E}_{\boldsymbol{p}} \left[h(\boldsymbol{x}, \boldsymbol{\xi}) \right],$$

where

$$\begin{aligned} h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) = \min_{\boldsymbol{y}} \; & \boldsymbol{d}_\omega^\top \boldsymbol{y} \\ s.t. \quad & \boldsymbol{q}_\omega^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{L}_\omega \boldsymbol{y} \geq \boldsymbol{c}_\omega^\top \boldsymbol{x}, \\ & \boldsymbol{D}_\omega \boldsymbol{y} \geq \boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega, \\ & \boldsymbol{y} \geq \boldsymbol{0}, \end{aligned} \tag{43}$$

or

$$\begin{aligned} h(\boldsymbol{x}, \boldsymbol{\xi}_\omega) = \min_{\boldsymbol{y}} \; & \boldsymbol{c}_\omega^\top \boldsymbol{x} + \boldsymbol{q}_\omega^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{L}_\omega \boldsymbol{y} \\ s.t. \quad & \boldsymbol{D}_\omega \boldsymbol{y} \geq \boldsymbol{B}_\omega \boldsymbol{x} + \boldsymbol{b}_\omega, \\ & \boldsymbol{y} \geq \boldsymbol{0}. \end{aligned}$$

Note that $G_\omega(\boldsymbol{x}, \boldsymbol{\lambda})$, defined in (29), has the same structure as (43), with bilinear stochasticity on the left-hand side. We skip the details of the modified algorithm for brevity.

4 Numerical Experiments

In this section, we consider a problem with a nonconvex recourse and a generic decision-dependent ambiguity set to illustrate the efficacy of disjunctive cuts. We provide computational comparative results on the performance of Algorithm 5 and solving the extensive formulation using a commercial nonconvex solver.

We note that as Algorithms 2 and 3 are developed using Benders'-type optimality cuts, we do not present computational results for them. Not surprisingly and as extensively reported in the literature, see, e.g., [56], our preliminary computational experiments showed the superiority of the decomposition-based cutting plane algorithm over solving the extensive formulation using a commercial nonconvex solver.

4.1 Multiproduct Newsvendor Problem with Price-Dependent Demand

For numerical experiments, we considered joint stocking and pricing decisions for a multiproduct newsvendor problem with uncertain price-dependent demand. Let n denote the number of products. For each $i \in [n]$, suppose that c_i denotes the per unit purchasing cost, s_i denotes the per unit salvage price, and b_i denotes the per unit back-order cost. For a fixed order quantity $q_i \in \mathbb{R}$, a fixed per unit selling price $r_i \in \mathbb{R}$, and demand $\xi_i \in \mathbb{R}$, cost function $h_i(q_i, r_i, \xi_i)$ is defined as:

$$\begin{aligned} h_i(q_i, r_i, \xi_i) &= c_i q_i - r_i \min\{q_i, \xi_i\} - s_i(q_i - \xi_i)_+ + b_i(\xi_i - q_i)_+ \\ &= (c_i - s_i)(q_i - \xi_i)_+ + (b_i - c_i + r_i)(\xi_i - q_i)_+ - \xi_i(r_i - c_i). \end{aligned}$$

Equivalently, we have

$$\begin{aligned} h_i(q_i, r_i, \xi_i) = \min_{y_i^+, y_i^-} \;& (c_i - s_i)y_i^+ + (b_i - c_i + r_i)y_i^- - \xi_i(r_i - c_i) \\ \text{s.t.} \;\;& y_i^+ - y_i^- \geq q_i - \xi_i, \\ & y_i^+, y_i^- \geq 0. \end{aligned} \tag{44}$$

Let $x_i = (q_i, r_i)$, $i \in [n]$, and $h(\boldsymbol{x}, \boldsymbol{\xi}) = \sum_{i \in [n]} h_i(x_i, \xi_i)$. Moreover, let $\mathcal{X} = \{\boldsymbol{x} = (\boldsymbol{q}, \boldsymbol{r}) : \sum_{i \in [n]} c_i q_i \leq d,\; q_i \geq 0,\; q_i \in \mathbb{Z},\; \underline{r}_i \leq r_i \leq \overline{r}_i,\; i \in [n]\}$. We assume that for $i \in [n]$, (i) $c_i - s_i > 0$, (ii) $c < \underline{r}_i$, and (iii) $b_i + \underline{r}_i > c_i$. We note that (i) and (iii) ensure that the critical ratio $0 < \beta_i = \frac{c_i - s_i}{b_i - s_i + \underline{r}_i} < 1$ is well defined. Hence, for any fixed price $\underline{r}_i \leq r_i \leq \overline{r}_i$ and without having constraint $\sum_{i \in [n]} c_i q_i \leq d$, the optimal order quantity q_i will be the left β_i-quantile with respect to the distribution of ξ_i (see e.g., [50]).

We formulate a DRO problem in the form of (1), where $\mathcal{P}(\boldsymbol{x})$ is a price-dependent ambiguity set in the form of (7), and

$$\mu_{0,i}(\boldsymbol{r}) = \overline{\mu}_i\big(1 + \sum_{j \in [n]} u_{ij}^{\mu} r_j\big), \tag{45}$$

$$\sigma_{0,i}(\boldsymbol{r})^2 = \overline{\sigma}_i^2\big(1 + \sum_{j \in [n]} u_{ij}^{\sigma} r_j\big), \tag{46}$$

$$\mu_{0,i}(\boldsymbol{r})^2 = \overline{\mu}_i^2\big(1 + 2\sum_{j \in [n]} u_{ij}^{\mu} r_j + \sum_{j \in [n]} (u_{ij}^{\mu})^2 r_j^2 + 2\sum_{j \in [n]} \sum_{j' \in [j]} u_{ij}^{\mu} u_{ij'}^{\mu} r_j r_{j'}\big),$$

for $i \in [n]$. Here, $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\sigma}}$ indicate the vector of empirical mean and standard deviation of the random demand $\boldsymbol{\xi}$. We assume that $u_{ii}^{\mu} < 0$, $i \in [n]$, implying that an increase in the price of product i leads to a decrease in the average demand for product i, ξ_i. Moreover, for $i \in [n]$, u_{ij}^{μ} and u_{ij}^{σ}, $i \neq j \in [n]$, capture the impact of other products' price on the mean and standard deviation of the demand for product i, ξ_i; emphasizing that the products may be substitutable. We refer to the described ambiguity set as `DD-A`. We also consider two other ambiguity sets in the form of (7) with increasingly further relaxation of price-dependency as follows:

- `DD-B`: $\mu_{0,i}(\boldsymbol{r})$ is defined as (45) in the first set of constraints in (7), and $\mu_{0,i}(\boldsymbol{r}) = \overline{\mu}_i$ in the second set of constraints in (7). Moreover, $\sigma_{0,i}(\boldsymbol{r})^2$ is defined as (46).
- `DD-C`: Same as `DD-B` but $\sigma_{0,i}(\boldsymbol{r})^2 = \overline{\sigma}_i^2$.

Several points are in order. First, (44) is in the form of problem (2) with an objective function in the form of (5), yielding a nonconvex recourse function. Second, in light of Lemma 3, we have $\mathcal{L} = \{\boldsymbol{\lambda} : 0 \leq \lambda_j \leq \max\{(\overline{s} + \overline{r})(\overline{q} - \underline{\xi}), \overline{b}(\overline{\xi} - \underline{q})\},\; j \in [s]\}$, where $\overline{s} = \max_{i \in [n]} s_i$, $\overline{r} = \max_{i \in [n]} \overline{r}_i$, $\overline{b} = \max_{i \in [n]} b_i$, $\overline{\xi} = \max_{i \in [n], \omega \in [N]} \xi_i^{\omega}$, $\underline{\xi} = \min_{i \in [n], \omega \in [N]} \xi_i^{\omega}$, $\overline{q} = \max_{i \in [n]} d/c_i$, and $\underline{q} = \min_{i \in [n]} d/c_i$. Third, in light of the discussion in Section 3.3, we have $0 \leq y_i^+ \leq q_i$ and $0 \leq y_i^- \leq \xi_i$ in problem (44). Finally, an optimal solution of the Lagrangian dual function in the form of (33) only depends on y_i^-.

4.2 Experiment Design

To conduct experiments, we first generated $\boldsymbol{t} = (\boldsymbol{c}^\top, \boldsymbol{s}^\top, \boldsymbol{b}^\top, \underline{\boldsymbol{r}}^\top, \overline{\boldsymbol{r}}^\top)^\top$ for a multiproduct newsvendor problem. We generated this information in a nested way so that a problem with a larger number of products would contain all the information for a problem with a smaller number of products. We set $u^{\mu}_{ij} = \exp(-\|t_i - t_j\|)$ for $i \neq j$, $u^{\mu}_{ii} = -0.3$, $i \in [n]$, and $u^{\mu}_{ij} = \exp(-\|t_i - t_j\|/2)$, $i, j \in [n]$. We generated realizations of the random demand ξ_i, $i \in [n]$, independently from a folded normal distribution with mean 6 and standard deviation 2. Again, these realizations were generated in a nested way. We present the computational results for $n \in \{1, 2, 3\}$ products, hyperparameters $\tau_1 = \underline{\tau}_2 = 0$, $\overline{\tau}_2 \in \{1, 2\}$ for the ambiguity set, and $N \in \{100, 200, 500, 1000, 2000, 3000, 5000, 10000, 20000\}$ scenarios. We report the average results over five training sets $\{\xi^{\omega}_i : i \in [n], \omega \in [N]\}$. Observe from (7) that $\tau_1 = 0$ and $\overline{\tau}_2 = 1$ enforce no deviation on the empirical first-order moment and no upper deviation on the empirical second-order moment, respectively. Also, $\underline{\tau}_1 = 0$ sets the lower bound on the second-order moment to zero.

4.3 Computational Results

In this section, we compare the computational performance of the decomposition-based disjunctive cutting plane algorithm, proposed in Algorithm 5, denoted as `DECOMPOSED`, with solving the MINLP deterministic equivalent formulation, presented in (13), using an off-the-shelf nonconvex solver. We denote this as `DEF`. We implemented `DECOMPOSED` and `DEF` in Python and used GUROBI 9.1.2 as a nonconvex solver. All experiments were performed on a Linux Ubuntu 20.04 environment using one single core of a PC with an Intel Core i7-9700 3.00 GHz processor and 32.00 GB of RAM, with a time limit of 3600 seconds.

Tables 1–3 reports the average computational results (over five training sets) to solve problem (1) with the ambiguity sets `DD-A` to `DD-C`, respectively. The values under column "Gap (%)" show the average gap for instances that could be solved optimally, and in parentheses, it shows the average gap for instances that could not be solved optimally within the time limit. Also, the values under column "Time (s)" show the average time (in seconds) for instances that could be solved optimally, and in parentheses, it shows the number of instances (out of five) that could not be solved optimally within the time limit.

Observe from Table 1 that `DECOMPOSED` found an optimal solution within the time limit for all instances (for three instances, an optimal solution was obtained when the algorithm stopped after finishing a current iteration beyond the time limit). Whereas `DEF` stopped with a nonzero optimality gap in some cases, on average between 8.64-28.58 %. In addition, for instances that could be solved optimally within the time limit with both approaches, `DECOMPOSED` often found an optimal solution with less computational effort (in seconds). We especially observe that `DEF` generally had a higher average computational time and optimality gap for instances with more scenarios and products. Moreover, an increase of $\overline{\tau}_2$ from 1 to 2—a wider range on the second-order moment—led to generally easier problems to solve for both `DECOMPOSED` and `DEF`. Using `DD-B` and partially relaxing the decision-dependency imposed in `DD-A` resulted in generally easier problems for both `DECOMPOSED` and `DEF` (Table 2). In particular, `DECOMPOSED` and `DEF` could obtain an optimal solution within the time limit in all instances, except for one instance for each. We still observe that `DEF` generally had a higher average computational time than `DECOMPOSED`, especially for instances with more scenarios and products. However, unlike the results with the ambiguity set `DD-A`, an increase of $\overline{\tau}_2$ from 1 to 2 led to generally more difficult problems to solve for both `DECOMPOSED` and `DEF`. Further relaxing the ambiguity set by using `DD-C` made the resulting DRO problem easier to solve for both `DECOMPOSED` and `DEF` (Table 3). Using this model resulted in similar

Table 1: Comparison of DEF and DECOMPOSED for problem (1) with ambiguity set DD-A and $\tau_1 = \underline{\tau}_2 = 0$.

			DECOMPOSED		DEF	
$\overline{\tau}_2$	n	N	Gap (%)	Time (s)	Gap (%)	Time (s)
1.0	1	100	0.0	0.07	0.0	0.05
		200	0.0	0.34	0.0	0.1
		500	0.0	0.34	0.0	0.22
		1000	0.0	0.77	0.0	0.71
		2000	0.0	1.46	0.0	1.68
		3000	0.0	2.21	0.01	2.18
		5000	0.0	2.57	0.01	7.37
	2	100	0.0	1.81	0.01	0.98
		200	0.0	6.53	0.01	25.04
		500	0.0	17.78	0.01	76.06
		1000	0.0	80.41	0.01	51.84
		2000	0.0	225.8	0.01	136.89
		3000	0.0	254.1	0.01	281.3
		5000	0.0	589.7	0.01	632.25
	3	100	0.0	78.66	0.13	28.51
		200	0.0	359.03	0.03	60.98
		500	0.0	74.54	0.01	469.85
		1000	0.0	234.15	0.02	1632.64
		2000	0.0	554.82	0.01 (21.15)	1825.09 (3)
		3000	0.0	929.31	- (8.64)	- (5)
		5000	0.0 (0.0)	1469.6 (2)	- (27.79)	- (5)
2.0	1	100	0.0	0.07	0.0	0.04
		200	0.0	0.35	0.0	0.11
		500	0.0	0.32	0.0	0.2
		1000	0.0	0.61	0.0	0.5
		2000	0.0	1.35	0.0	1.16
		3000	0.0	1.73	0.01	1.73
		5000	0.0	2.1	0.0	3.51
	2	100	0.0	1.71	0.01	0.83
		200	0.0	6.95	0.01	2.48
		500	0.0	16.99	0.01	8.62
		1000	0.0	35.47	0.01	27.06
		2000	0.0	115.7	0.01	101.71
		3000	0.0	383.49	0.02	283.51
		5000	0.0	281.05	0.02	414.13
	3	100	0.0	74.26	0.68	28.37
		200	0.0	309.6	0.32	54.13
		500	0.0	72.03	0.05	503.4
		1000	0.0	201.25	0.07	1654.14
		2000	0.0	384.08	0.06 (22.12)	1139.24 (2)
		3000	0.0	768.88	0.09 (17.67)	2084.86 (3)
		5000	0.0 (0.0)	1351.46 (1)	- (28.58)	- (5)

trends as those observed using DD-B.

5 Conclusion

This paper studied a two-stage stochastic mixed-integer program with continuous recourse. We assumed that the probability distribution of random parameters is unknown and depends on decisions. We thus investigated a distributionally robust approach to this problem, where the distributional ambiguity is modeled with a polyhedral decision-dependent ambiguity set. We considered cases where the recourse function and the ambiguity set are either generic or have a special

Table 2: Comparison of `DEF` and `DECOMPOSED` for problem (1) with ambiguity set `DD-B` and $\tau_1 = \underline{\tau}_2 = 0$.

			`DECOMPOSED`		`DEF`	
$\overline{\tau}_2$	n	N	Gap (%)	Time (s)	Gap (%)	Time (s)
1.0	1	100	0.0	0.05	0.0	0.08
		200	0.0	0.08	0.0	0.17
		500	0.0	0.18	0.0	0.54
		1000	0.0	0.43	0.01	1.23
		2000	0.0	1.19	0.0	1.85
		3000	0.0	1.76	0.01	3.31
		5000	0.0	2.17	0.01	10.58
	2	100	0.0	0.47	0.01	0.45
		200	0.0	1.64	0.01	0.78
		500	0.0	3.97	0.01	2.37
		1000	0.0	9.33	0.01	6.33
		2000	0.0	25.92	0.01	42.08
		3000	0.0	36.29	0.01	71.75
		5000	0.0	72.08	0.01	156.08
	3	100	0.0	0.7	0.01	0.85
		200	0.0	1.67	0.01	1.18
		500	0.0	5.62	0.01	4.45
		1000	0.0	11.01	0.01	12.59
		2000	0.0	18.97	0.01	39.62
		3000	0.0	28.06	0.01	87.55
		5000	0.0	63.78	0.01	101.83
2.0	1	100	0.0	0.07	0.0	0.03
		200	0.0	0.41	0.0	0.08
		500	0.0	0.24	0.0	0.19
		1000	0.0	0.54	0.0	0.34
		2000	0.0	1.11	0.0	0.52
		3000	0.0	1.78	0.0	0.71
		5000	0.0	2.08	0.0	1.99
	2	100	0.0	9.44	0.01	1.23
		200	0.0	140.88	0.01	2.67
		500	0.0	464.54	0.01	7.32
		1000	0.0	22.26	0.01	36.82
		2000	0.0	875.43	0.01	89.5
		3000	0.0	121.92	0.01	233.59
		5000	0.0	240.46	0.01	355.71
	3	100	0.0	130.65	0.01	23.06
		200	0.0	4.42	0.01	29.16
		500	0.0	598.03	0.01	112.07
		1000	0.0 (0.12)	24.21 (1)	0.01	267.23
		2000	0.0	51.0	0.01	686.06
		3000	0.0	76.77	0.01	1156.22
		5000	0.0	123.76	0.01 (0.02)	1902.47 (1)

convex/nonconvex structure. We reformulated the resulting problem as a nonconvex two-stage stochastic mixed-integer program. We proposed finitely-convergent decomposition-based cutting plane algorithms to obtain an ϵ-optimal solution to the resulting problems. The proposed algorithm for the case that the recourse function is nonconvex with a bilinear objective function is of independent interest to solve two-stage stochastic programs with a random decision-dependent recourse matrix (i.e., bilinear stochasticity on the left-hand side). We illustrated the efficacy of the proposed algorithm when the recourse function is nonconvex on joint pricing and stocking decisions for a multiproduct newsvendor problem with price-dependent demand.

This paper focused on a distributionally robust optimization problem with a finite sample

Table 3: Comparison of DEF and DECOMPOSED for problem (1) with ambiguity set DD-C and $\tau_1 = \underline{\tau}_2 = 0$.

			DECOMPOSED		DEF	
$\overline{\tau}_2$	n	N	Gap (%)	Time (s)	Gap (%)	Time (s)
1.0	1	100	0.0	0.03	0.0	0.07
		200	0.0	0.04	0.0	0.13
		500	0.0	0.17	0.0	0.48
		1000	0.0	0.22	0.0	1.03
		2000	0.0	0.43	0.01	2.43
		3000	0.0	0.93	0.01	3.67
		5000	0.0	0.97	0.01	9.46
	2	100	0.0	0.12	0.01	0.24
		200	0.0	0.3	0.0	0.39
		500	0.0	0.74	0.01	0.92
		1000	0.0	1.57	0.0	2.61
		2000	0.0	3.13	0.0	8.38
		3000	0.0	3.07	0.0	16.66
		5000	0.0	5.26	0.0	38.67
	3	100	0.0	0.51	0.0	0.14
		200	0.0	1.14	0.0	0.32
		500	0.0	2.74	0.0	1.42
		1000	0.0	5.57	0.0	3.47
		2000	0.0	9.67	0.0	9.26
		3000	0.0	14.92	0.0	20.78
		5000	0.0	25.55	0.0	35.8
2.0	1	100	0.0	0.06	0.0	0.06
		200	0.0	0.58	0.0	0.08
		500	0.0	0.29	0.0	0.27
		1000	0.0	0.58	0.0	0.78
		2000	0.0	1.11	0.0	0.62
		3000	0.0	1.67	0.0	0.92
		5000	0.0	2.26	0.0	1.82
	2	100	0.0	6.17	0.01	0.65
		200	0.0	14.31	0.01	1.26
		500	0.0	108.12	0.01	7.57
		1000	0.0	158.88	0.01	22.52
		2000	0.0	189.79	0.01	81.37
		3000	0.0	286.61	0.01	120.42
		5000	0.0	72.93	0.01	370.42
	3	100	0.0	0.5	0.0	0.99
		200	0.0	1.01	0.01	1.32
		500	0.0	3.1	0.01	4.13
		1000	0.0	5.95	0.01	12.48
		2000	0.0	9.43	0.01	38.79
		3000	0.0	15.98	0.01	86.06
		5000	0.0	30.47	0.0	322.41

space and a polyhedral ambiguity set. Using Lagrangian/linear programming duality, we obtained reformulations that serve as a basis for the proposed decomposition-based cutting plane algorithms. Future work includes investigating the case that the sample space is infinite. As in the DRO literature with a decision-independent ambiguity set, more generalized forms of duality, e.g., conic duality, are expected to be needed for reformulation. On the other hand, the reformulated problem is expected to be a semi-infinite program. Especially for the case that the recourse function is nonconvex, it would be interesting to explore how the proposed disjunctive cutting plane algorithm may be extended. Another direction for future research is to investigate stochastic programs with probabilistic constraints.

Acknowledgements

The authors gratefully acknowledge the support of the Office of Naval Research through grant N00014-18-1-2097-P00001.

References

[1] Ahmed, S. (2000). *Strategic planning under uncertainty: Stochastic integer programming approaches.* Ph.D. dissertation, University of Illinois at Urbana-Champaign.

[2] Balas, E. (1998). Disjunctive programming: Properties of the convex hull of feasible points. *Discrete Appl. Math. 89*(1-3), 3–44.

[3] Bansal, M. and S. Mehrotra (2019). On solving two-stage distributionally robust disjunctive programs with a general ambiguity set. *Eur. J. Oper. Res. 279*(2), 296–307.

[4] Basciftci, B., S. Ahmed, and N. Gebraeel (2020). Data-driven maintenance and operations scheduling in power systems under decision-dependent uncertainty. *IISE Transactions 52*(6), 589–602.

[5] Basciftci, B., S. Ahmed, and S. Shen (2021). Distributionally robust facility location problem under decision-dependent stochastic demand. *Eur. J. Oper. Res. 292*(2), 548–561.

[6] Bayraksan, G. and D. K. Love (2015). Data-driven stochastic programming using phi-divergences. In *The Operations Research Revolution*, pp. 1–19. INFORMS TutORials in Operations Research.

[7] Ben-Tal, A., D. den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Sci. 59*(2), 341–357.

[8] Ben-Tal, A., L. El Ghaoui, and A. Nemirovski (2009). *Robust Optimization.* Princeton University Press.

[9] Benders, J. F. (1962, Dec). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik 4*(1), 238–252.

[10] Benson, H. P. (1995). Concave minimization: theory, applications and algorithms. In R. Horst and P. M. Pardalos (Eds.), *Handbook of Global Optimization*, Volume 2 of *Nonconvex Optimization and Its Applications*, pp. 43–148. Springer.

[11] Bertsimas, D., D. B. Brown, and C. Caramanis (2011). Theory and applications of robust optimization. *SIAM Rev. 53*(3), 464–501.

[12] Bertsimas, D. and I. Popescu (2005). Optimal inequalities in probability theory: A convex optimization approach. *SIAM J. Optim. 15*(3), 780–804.

[13] Bienstock, D., C. Chen, and G. Munoz (2020). Outer-product-free sets for polynomial optimization and oracle-based cuts. *Math. Program. 183*(1), 105–148.

[14] Birge, J. R. and F. Louveaux (2011). *Introduction to stochastic programming* (2nd ed.). Springer Science & Business Media.

[15] Blanchet, J. and K. Murthy (2019). Quantifying distributional model risk via optimal transport. *Math. Oper. Res. 44*(2), 565–600.

[16] Burer, S. and A. N. Letchford (2012). Non-convex mixed-integer nonlinear programming: A survey. *Surv. Oper. Res. Management Sci. 17*(2), 97–106.

[17] Colvin, M. and C. T. Maravelias (2008). A stochastic programming approach for clinical trial planning in new drug development. *Comput. Chem. Eng 32*(11), 2626–2642.

[18] Davarnia, D., J.-P. P. Richard, and M. Tawarmalani (2017). Simultaneous convexification of bilinear functions over polytopes with application to network interdiction. *SIAM J. Optim. 27*(3), 1801–1833.

[19] Delage, E. and Y. Ye (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res. 58*(3), 595–612.

[20] Dentcheva, D. and A. Ruszczyński (2020). Risk forms: representation, disintegration, and application to partially observable two-stage systems. *Math. Program. 181*(2), 297–317.

[21] Dupačová, J. (2006). Optimization under exogenous and endogenous uncertainty. In L. L (Ed.), *Proceedings of MME06*. University of West Bohemia in Pilsen.

[22] Duque, D., S. Mehrotra, and D. P. Morton (2022). Distributionally robust two-stage stochastic programming. *SIAM J. Optim. 32*(3), 1499–1522.

[23] Gamboa, C. A., D. M. Valladão, A. Street, and T. Homem-de Mello (2021). Decomposition methods for wasserstein-based data-driven distributionally robust problems. *Oper. Res. Lett. 49*(5), 696–702.

[24] Gao, R. and A. Kleywegt (2023). Distributionally robust stochastic optimization with Wasserstein distance. *Math. Oper. Res. 48*(2), 603–655.

[25] Goel, V. and I. E. Grossmann (2006). A class of stochastic programs with decision dependent uncertainty. *Math. Program. 108*(2), 355–394.

[26] Hellemo, L., P. I. Barton, and A. Tomasgard (2018). Decision-dependent probabilities in stochastic programs with recourse. *Comput. Management Sci. 15*(3), 369–395.

[27] Horst, R. and H. Tuy (1996). *Global optimization: Deterministic approaches* (3rd ed.). Springer Science & Business Media.

[28] Hu, J., M. Bansal, and S. Mehrotra (2018). Robust decision making using a general utility set. *Eur. J. Oper. Res. 269*(2), 699–714.

[29] Hu, J., J. Li, and S. Mehrotra (2019). A data-driven functionally robust approach for simultaneous pricing and order quantity decisions with unknown demand function. *Oper. Res. 67*(6), 1564–1585.

[30] Ji, R. and M. A. Lejeune (2021). Data-driven optimization of reward-risk ratio measures. *INFORMS J. Comput. 33*(3), 1120–1137.

[31] Jiang, R. and Y. Guan (2018). Risk-averse two-stage stochastic program with distributional ambiguity. *Oper. Res. 66*(5), 1390–1405.

[32] Jonsbråten, T. W., R. J. Wets, and D. L. Woodruff (1998). A class of stochastic programs with decision dependent random elements. *Annals of Oper. Res. 82*, 83–106.

[33] Kuhn, D., P. M. Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh (2019). Wasserstein distributionally robust optimization: Theory and applications in machine learning. In *Operations Research & Management Science in the Age of Analytics*, pp. 130–166. INFORMS TutORials in Operations Research.

[34] Luo, F. (2020). A distributionally-robust service center location problem with decision dependent demand induced from a maximum attraction principle. `arXiv preprint arXiv:2011.12514[math.OC]`.

[35] Luo, F. and S. Mehrotra (2019). Decomposition algorithm for distributionally robust optimization using Wasserstein metric with an application to a class of regression models. *Eur. J. Oper. Res. 278*(1), 20–35.

[36] Luo, F. and S. Mehrotra (2020). Distributionally robust optimization with decision dependent ambiguity sets. *Optim. Lett. 14*(8), 2565–2594.

[37] Luo, F. and S. Mehrotra (2024). Service center location problems with decision dependent utilities and a pandemic case study. *Nav. Res. Logist. 71*(1), 3–26.

[38] Mehrotra, S. and D. Papp (2014). A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization. *SIAM J. Optim. 24*(4), 1670–1697.

[39] Meyer, R. (1970). The validity of a family of optimization methods. *SIAM J. Control 8*(1), 41–54.

[40] Mohajerin Esfahani, P. and D. Kuhn (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Prog. 171*(1-2), 115–166.

[41] Nohadani, O. and A. Roy (2017). Robust optimization with time-dependent uncertainty in radiation therapy. *IISE Trans. Healthc. Syst. Eng. 7*(2), 81–92.

[42] Nohadani, O. and K. Sharma (2018). Optimization under decision-dependent uncertainty. *SIAM J. Optim. 28*(2), 1773–1795.

[43] Noyan, N., G. Rudolf, and M. Lejeune (2022). Distributionally robust optimization under a decision-dependent ambiguity set with applications to machine scheduling and humanitarian logistics. *INFORMS J. Comput. 34*(2), 729–751.

[44] Peeta, S., F. S. Salman, D. Gunnec, and K. Viswanath (2010). Pre-disaster investment decisions for strengthening a highway network. *Comput. Oper. Res. 37*(10), 1708–1719.

[45] Perakis, G. and G. Roels (2008). Regret in the newsvendor model with partial information. *Oper. Res. 56*(1), 188–203.

[46] Popescu, I. (2005). A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Math. Oper. Res. 30*(3), 632–657.

[47] Porembski, M. (2001). Finitely convergent cutting planes for concave minimization. *J. Global Optim. 20*, 109–132.

[48] Porembski, M. (2002). Cone adaptation strategies for a finite and exact cutting plane algorithm for concave minimization. *J. Global Optim. 24*, 89–107.

[49] Powell, W. B. and H. Topaloglu (2003). Stochastic programming in transportation and logistics. In A. Ruszczyński and A. Shapiro (Eds.), *Stochastic Programming*, Volume 10 of *Handbooks in Operations Research and Management Science*, pp. 555–635. Elsevier.

[50] Rahimian, H., G. Bayraksan, and T. Homem-de Mello (2019a). Controlling risk and demand ambiguity in newsvendor models. *Eur. J. Oper. Res. 279*(3), 854–868.

[51] Rahimian, H., G. Bayraksan, and T. Homem-de Mello (2019b). Identifying effective scenarios in distributionally robust stochastic programs with total variation distance. *Math. Program. 173*(1-2), 393–430.

[52] Rahimian, H., G. Bayraksan, and T. Homem-de Mello (2022). Effective scenarios in multistage distributionally robust optimization with a focus on total variation distance. *SIAM J. Optim. 32*(3), 1698–1727.

[53] Rahimian, H. and S. Mehrotra (2022a). A finitely convergent disjunctive cutting plane algorithm for bilinear programming. Optimization Online https://optimization-online.org/?p=16220.

[54] Rahimian, H. and S. Mehrotra (2022b). Frameworks and results in distributionally robust optimization. *Open J. Math. Optim 3.* article no. 4, 85 p. https//doi.org/10.5802/ojmo.15.

[55] Rahimian, H. and S. Mehrotra (2022c). Sequential convexification of a bilinear set. Optimization Online https://optimization-online.org/?p=16221.

[56] Rahmaniani, R., T. G. Crainic, M. Gendreau, and W. Rei (2017). The benders decomposition algorithm: A literature review. *Eur. J. Oper. Res. 259*(3), 801–817.

[57] Royset, J. O. and R. J.-B. Wets (2017). Variational theory for optimization under stochastic ambiguity. *SIAM J. Optim. 27*(2), 1118–1149.

[58] Ryoo, H. S. and N. V. Sahinidis (1995). Global optimization of nonconvex nlps and MINLPs with applications in process design. *Comput. Chem. Eng. 19*(5), 551–566.

[59] Salmeron, J., R. K. Wood, and D. P. Morton (2009). A stochastic program for optimizing military sealift subject to attack. *Mil. Oper. Res.*, 19–39.

[60] Saxena, A., P. Bonami, and J. Lee (2010). Convex relaxations of non-convex mixed integer quadratically constrained programs: extended formulations. *Math. Program. 124*(1-2), 383–411.

[61] Scala, N. M. and I. Howard, James P (Eds.) (2020). *Handbook of Military and Defense Operations Research.* Series in Operations Research. CRC Press.

[62] Scarf, H. (1958). A min-max solution of an inventory problem. In H. Scarf, K. Arrow, and S. Karlin (Eds.), *Studies in the mathematical theory of inventory and production*, pp. 201–209. Stanford University Press, Stanford, CA.

[63] Shapiro, A. (2017). Distributionally robust stochastic programming. *SIAM J. Optim. 27*(4), 2258–2275.

[64] Shapiro, A. and S. Ahmed (2004). On a class of minimax stochastic programs. *SIAM J. Optim. 14*(4), 1237–1249.

[65] Shapiro, A., D. Dentcheva, and A. Ruszczyński (2014). *Lectures on stochastic programming: modeling and theory* (2nd ed.). MPS-SIAM series on optimization. Society for Industrial and Applied Mathematics, Philadelphia, USA.

[66] Sherali, H. D. and A. Alameddine (1992). A new reformulation-linearization technique for bilinear programming problems. *J. Global Optim. 2*(4), 379–410.

[67] Sherali, H. D. and C. Shetty (1980). A finitely convergent algorithm for bilinear programming problems using polar cuts and disjunctive face cuts. *Math. Program. 19*(1), 14–31.

[68] Smith, E. M. and C. C. Pantelides (1997). Global optimisation of nonconvex MINLPs. *Comput. Chem. Eng. 21*, S791–S796.

[69] Van Slyke, R. M. and R. Wets (1969). L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM J. Appl. Math. 17*(4), 638–663.

[70] Vayanos, P., A. Georghiou, and H. Yu (2020). Robust optimization with decision-dependent information discovery. `arXiv preprint arXiv:2004.08490 [math.OC]`.

[71] Vayanos, P., D. Kuhn, and B. Rustem (2011). Decision rules for information discovery in multi-stage stochastic programming. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pp. 7368–7373.

[72] Wets, R. J.-B. (1985). On the continuity of the value of a linear program and of related polyhedral-valued multifunctions. In R. W. Cottle (Ed.), *Math. Program. Essays in Honor of George B. Dantzig Part I*, Math. Program. Studies, Chapter 2, pp. 14–29. Springer.

[73] Wiesemann, W., D. Kuhn, and M. Sim (2014). Distributionally robust convex optimization. *Oper. Res. 62*(6), 1358–1376.

[74] Yu, X. and S. Shen (2022). Multistage distributionally robust mixed-integer programming with decision-dependent moment-based ambiguity sets. *Math. Program. 196*(1-2), 1025–1064.

[75] Zhao, C. and Y. Guan (2018). Data-driven risk-averse stochastic optimization with Wasserstein metric. *Oper. Res. Lett. 46*(2), 262–267.