Adjustable Distributionally Robust Optimization with Infinitely Constrained Ambiguity Sets

Haolin Ruan
School of Data Science, City University of Hong Kong, Kowloon Tong, Hong Kong
haolin.ruan@my.cityu.edu.hk

Zhi Chen
Department of Management Sciences, College of Business, City University of Hong Kong, Kowloon Tong, Hong Kong
zhi.chen@cityu.edu.hk

Chin Pang Ho
School of Data Science, City University of Hong Kong, Kowloon Tong, Hong Kong
clint.ho@cityu.edu.hk

We study adjustable distributionally robust optimization problems where their ambiguity sets can potentially encompass an infinite number of expectation constraints. Although such an ambiguity set has great modeling flexibility in characterizing uncertain probability distributions, the corresponding adjustable problems remain computationally intractable and challenging. To overcome this issue, we propose a greedy improvement procedure that consists of solving, via the (extended) linear decision rule approximation, a sequence of tractable subproblems—each of which considers a relaxed and finitely constrained ambiguity set that is also iteratively tightened to the infinitely constrained one. Through three numerical studies of adjustable distributionally robust optimization models, we show that our approach can yield improved solutions in a systematic way for both two-stage and multi-stage problems.

Key words: adjustable optimization, distributionally robust optimization, infinitely constrained ambiguity set, linear decision rule
History: March 12, 2022

1. Introduction

Distributionally robust optimization is one of the most popular approaches for addressing decision-making problems (*e.g.*, operations management problems; see Lu and Shen 2021) in the face of uncertainty. In distributionally robust optimization models, uncertainty is modeled as a random variable that is governed by an unknown probability distribution residing in an ambiguity set—a family of distributions that share some common distributional information, and the decision maker seeks solutions that are immune against all possible candidates from within the ambiguity set. The introduction of ambiguity set leads to greater modeling flexibility and allows a modeler to encode, in a unified fashion, rich types of information about the uncertainty, such as support, moments, descriptive statistics, and even structure information (see, *e.g.*, Ben-Tal and Nemirovski 1998, Bertsimas and Sim 2004, Chen et al. 2021b, Delage and Ye 2010, Li et al. 2019, Wiesemann et al. 2014). Recent growing interest also lies in designing ambiguity sets whose member distributions are not far away from a reference distribution, where the proximity is measured by some statistical distances including the Wasserstein metric (Bertsimas et al. 2021, Blanchet and Murthy 2019, Gao and Kleywegt 2016, Mohajerin Esfahani and Kuhn 2018) as well as ϕ-divergence (Ben-Tal et al. 2013, Duchi et al. 2021, Wang et al. 2016). There are also novel perspectives on the design of an ambiguity set, such as incorporating Bayesian approach (Gupta 2019) and hypothesis testing (Bertsimas et al. 2018a,b).

Adjustable distributionally robust optimization is an important class of models for multi-stage decision-making problems, where in each stage, new wait-and-see decisions are adjustable to revealed uncertain parameters.Applications of adjustable distributionally robust optimization appear in a rich variety of domains such as network design (Atamtürk and Zhang 2007), transportation (He et al. 2020), inventory management (Bandi et al. 2019), and energy systems (Lorca et al. 2016), just to name a few. On the technical side, extending to adjustable distributionally robust optimization remains challenging. The major computational difficulty is that adjustable decisions are typically modeled as infinite-dimensional functions (or, decision rules) of the revealed uncertainty, and they are hard to be optimized. Ben-Tal et al. (2004) show that the adjustable robust counterparts of uncertain linear programs are already computationally intractable in most cases, while Shapiro and Nemirovski (2005) argue that multi-stage problems are generally computationally intractable even when only medium-quality solutions are sought. Another possible issue of adjustable distributionally robust optimization is scalability—the computation time of obtaining its exact solution often easily becomes intolerable as the problem size grows (see, *e.g.*, Zhen et al. 2018). For the sake of tractability and scalability, Ben-Tal et al. (2004) suggest to restrict the admissible functional forms of adjustable decisions to affine functions of the uncertainty—an approach that is known as linear decision rule (LDR)[1] approximation and has been discussed in the early literature of stochastic programming (Garstka and Wets 1974). The effectiveness and computational attractiveness presented by Ben-Tal et al. (2004) rekindle the interest in adopting this approach in distributionally robust optimization. Many works follow this spirit and further refine adjustable decisions (*i*) to be piecewise affine (see, *e.g.*, Chen et al. 2007, Goh and Sim 2010) and (*ii*) to further depend on auxiliary random variables that arise from the corresponding lifted ambiguity set (see, *e.g.*, Bertsimas et al. 2019, Chen and Zhang 2009, Chen et al. 2020). Certainly there are concerns about the general suboptimality of LDR approximation (*e.g.*, Garstka and Wets

[1] Throughout this paper we use the term "linear decision rule", which is also called affine decision rule. Both are commonly used in the literature.

1974, Bertsimas and Goyal 2012), it is also worth mentioning that the LDR approximation could perform reasonably well in some applications (*e.g.*, Ben-Tal et al. 2005) and sometimes can even be optimal (see, for instance, Anderson and Moore 2007, Bertsimas et al. 2010, Gounaris et al. 2013, and Iancu et al. 2013). The following quote from Shapiro and Nemirovski (2005) clarifies the rationale for considering LDR approximation: *"The only reason for restricting ourselves with affine decision rules stems from the desire to end up with a computationally tractable problem."* We refer to Georghiou et al. (2021) for a comprehensive review on linear decision rules in distributionally robust optimization.

In this paper, we focus on the class of infinitely constrained ambiguity sets proposed by Chen et al. (2019), in which the number of expectation constraints could be infinite. This class of ambiguity sets can incorporate rich distributional information of the uncertainty, including, among other things, the stochastic dominance, mean-dispersion, fourth moment, and entropic dominance. Due to the infinite number of expectation constrains, even the corresponding static distributionally robust optimization problems may not admit tractable reformulations. Motivated by the wide applications of adjustable distributional robust optimization as well as the modeling power of the infinitely constrained ambiguity set, we focus on adjustable problems with such an ambiguity set. In particular, we propose a greedy improvement procedure to obtain high-quality approximate solutions. Our main contributions may be summarized as follows.

(*i*) We introduce the infinitely constrained ambiguity set developed in Chen et al. (2019) to two-stage/multi-stage adjustable distributionally robust optimization problems. The infinitely constrained set possesses great modeling flexibility and advantages that include, for example, characterizing complete covariance information via second-order cone inequalities. In contrast, an ambiguity set with a finite number of expectation constraints based on the second-order cone would fail to achieve this.

(*ii*) To solve the corresponding adjustable distributionally robust optimization problems, we propose a greedy improvement procedure that solve a sequence of tractable subproblems. In each of these subproblems, we consider a relaxed ambiguity set with a finite subset of expectation constraints, and we adopt the extended LDR approximation where adjustable decisions depend on both primary random variables (*i.e.*, the uncertainty) and auxiliary random variables arising from the lifted ambiguity set. At each iteration, we identify violated expectation constraints to be included in the relaxed ambiguity set for a better extended LDR approximation.

(*iii*) In numerical experiments on three applications, we focus on adjustable distributionally robust optimization problems that consider complete covariance information. We show that our approach can, in a systematic way, adapt and improve from partial cross-moments (towards complete covariance) for two-stage/multi-stage problems, providing a positive answer to the question raised by Bertsimas et al. (2019, § 6) on "*how we can systematically adapt and improve the partial cross-moments ambiguity set*".

The remainder of this paper proceeds as follows. Section 2 introduces the two-stage adjustable optimization framework and Section 3 incorporates the extended LDR approximation. We propose in Section 4 the greedy procedure to iteratively tighten the extended LDR approximation. Section 5 extends to multi-stage problems. In Section 6, we showcase the encouraging performance of our proposed framework and solution approach via a newsvendor problem, a hospital quota allocation problem and an inventory control problem. Section 7 concludes our work.

Notation: We use boldface lowercase characters to represent vectors, which by default, are column vectors. An exception is that we denote the m-th row of a matrix $\boldsymbol{A}$ by a row vector $\boldsymbol{A}_{m:}$. Special vectors include $\mathbf{0}$, $\mathbf{1}$ and $\boldsymbol{e}_i$, which are vectors of all zeros, vectors of all ones and the i-th standard unit basis, respectively. The set of positive running indices up to N is $[N]=\{1,2,\dots,N\}$. A random variable $\tilde{\boldsymbol{z}}\sim\mathbb{P}$ is denoted with a tilde sign and is governed by $\mathbb{P}\in\mathcal{P}_0(\mathbb{R}^I)$, where $\mathcal{P}_0(\mathbb{R}^I)$ represents the set of all probability distributions on $\mathbb{R}^I$. Given a probability distribution $\mathbb{P}$, $\mathbb{E}_{\mathbb{P}}[\cdot]$ denotes the corresponding expectation. We say that a set is tractable conic representable if its membership can be described by finitely many linear/second-order cone/exponential cone constraints and, potentially, auxiliary variables. A function is tractable conic representable if its epigraph is.

2. Adjustable Distributionally Robust Optimization

Consider a two-stage adjustable optimization problem where the *here-and-now* decision $\boldsymbol{x}\in\mathbb{R}^N$ is chosen over the feasible set $\mathcal{X}$. The first-stage cost is deterministic and is given by $\boldsymbol{c}^\top\boldsymbol{x}$ for some $\boldsymbol{c}\in\mathbb{R}^N$. In progressing to the second stage, the random variable $\tilde{\boldsymbol{z}}\in\mathbb{R}^I$ with support $\mathcal{W}\subseteq\mathbb{R}^I$ is realized; thereafter, we could determine the cost incurred at the second stage. Similar to a typical stochastic programming model, for a given decision $\boldsymbol{x}$ and a realization $\boldsymbol{z}$, we evaluate the second-stage cost via a linear program that involves the adjustable *wait-and-see* decision $\boldsymbol{y}\in\mathbb{R}^L$:

$$f(\boldsymbol{x},\boldsymbol{z})=\begin{cases}\min\ \boldsymbol{d}^\top\boldsymbol{y}\\ \text{s.t.}\ \ \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}\geq\boldsymbol{b}(\boldsymbol{z})\\ \qquad \boldsymbol{y}\in\mathbb{R}^L.\end{cases} \tag{1}$$

Here, we adopt the popular factor-based model that assumes $\boldsymbol{A}$ and $\boldsymbol{b}$ are affine in $\boldsymbol{z}$. That is,

$$\boldsymbol{A}(\boldsymbol{z})=\boldsymbol{A}_0+\sum_{i\in[I]}\boldsymbol{A}_i z_i \quad \text{and} \quad \boldsymbol{b}(\boldsymbol{z})=\boldsymbol{b}_0+\sum_{i\in[I]}\boldsymbol{b}_i z_i$$

with $\boldsymbol{A}_i\in\mathbb{R}^{M\times N}$ and $\boldsymbol{b}_i\in\mathbb{R}^M$ for $i\in[I]\cup\{0\}$. The elements in both the recourse matrix $\boldsymbol{B}\in\mathbb{R}^{M\times L}$ and the vector of cost parameters $\boldsymbol{d}\in\mathbb{R}^L$ are constant, and this setting is referred to as fixed recourse in the terminology of stochastic programming. In general, problem (1) could be infeasible, as the recourse matrix can largely influence its feasibility; see, for example, in the case of complete recourse or relatively complete recourse whose definitions are given below.

Definition 1 (Complete recourse). The second-stage problem (1) has complete recourse if there exists $\boldsymbol{y} \in \mathbb{R}^L$ such that $\boldsymbol{B}\boldsymbol{y} > \boldsymbol{0}$.

Complete recourse is a strong sufficient condition that guarantees the feasibility of the second-stage problem for all $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{z} \in \mathbb{R}^I$. This implies that the second-stage cost $f(\boldsymbol{x}, \boldsymbol{z}) < +\infty$ for any $\boldsymbol{x}$ and $\boldsymbol{z}$. Typically, a weaker condition below is assumed in stochastic programming to ensure that the second-stage problem is feasible.

Definition 2 (Relatively complete recourse). The second-stage problem (1) has relatively complete recourse if and only if the problem is feasible for all $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{z} \in \mathcal{W}$.

On top of these two conditions, the following sufficiently expensive recourse condition is also often considered due to practical interest.

Definition 3 (Sufficiently expensive recourse). The second-stage problem (1) has sufficiently expensive recourse if the second-stage cost $f(\boldsymbol{x}, \boldsymbol{z}) > -\infty$ for all $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{z} \in \mathcal{W}$.

The following result reveals a relation between the cost vector $\boldsymbol{d}$ and the recourse matrix $\boldsymbol{B}$ under the sufficiently expensive recourse condition.

Lemma 1. *Under the sufficiently expensive recourse condition, the vector of cost parameters* $\boldsymbol{d}$ *is a non-negative linear combination of rows of the recourse matrix* $\boldsymbol{B}$.

Proof of Lemma 1. Suppose, on the contrary, that $\boldsymbol{d}$ is not a non-negative linear combination of $\{\boldsymbol{B}_{m:}\}_{m\in[M]}$, *i.e.*, the problem $\max_{\boldsymbol{p}}\{0 \mid \boldsymbol{B}^\top \boldsymbol{p} = \boldsymbol{d},\ \boldsymbol{p} \geq \boldsymbol{0}\}$ is infeasible. Then the dual problem $\min_{\boldsymbol{q}}\{\boldsymbol{d}^\top \boldsymbol{q} \mid \boldsymbol{B}\boldsymbol{q} \geq \boldsymbol{0}\}$ is unbounded because $\boldsymbol{q} = \boldsymbol{0}$ is always a feasible solution. Hence, there must be some $\boldsymbol{q}$ satisfying $\boldsymbol{d}^\top \boldsymbol{q} < 0$. This, however, contradicts to sufficiently expensive recourse that requires $f(\boldsymbol{x}, \boldsymbol{z})$ to be bounded from below. □

In this paper, to assure the feasibility and the boundedness of the second-stage problem, relatively complete recourse and sufficiently expensive recourse are always assumed in our framework. We are interested in solving the following adjustable distributionally robust optimization problem

$$\begin{aligned} \min\ & \boldsymbol{c}^\top \boldsymbol{x} + \rho(\boldsymbol{x}) \\ \text{s.t.}\ & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{2}$$

Here, the optimal here-and-now decision $\boldsymbol{x}$ minimizes the sum of the deterministic first-stage cost and its corresponding worst-case expected second-stage cost,

$$\rho(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})], \tag{3}$$

under an infinitely constrained ambiguity set $\mathcal{F}$ of the form

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^I\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{q}, \tilde{\boldsymbol{z}})] \leq h(\boldsymbol{q}) \quad \forall \boldsymbol{q} \in \mathcal{Q} \\ \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{W}] = 1 \end{array} \right\} \tag{4}$$

with $\boldsymbol{\mu} \in \mathbb{R}^I$, $\mathcal{Q} \subseteq \mathbb{R}^I$, $g : \mathcal{Q} \times \mathbb{R}^I \mapsto \mathbb{R}$ and $h : \mathcal{Q} \mapsto \mathbb{R}$. The support set $\mathcal{W}$ is non-empty, bounded, and tractable conic representable. For any given $\boldsymbol{q} \in \mathcal{Q}$, the function $g(\boldsymbol{q}, \boldsymbol{z})$ is tractable conic representable with respect to $\boldsymbol{z}$. We also assume that $\boldsymbol{\mu} \in \mathcal{W}$ and $g(\boldsymbol{q}, \boldsymbol{\mu}) \leq h(\boldsymbol{q})$ for all $\boldsymbol{q} \in \mathcal{Q}$. The inclusion of a possibly infinite number of expectation constraints grants $\mathcal{F}$ great modeling power. Chen et al. (2019) show that a generic distributionally robust optimization problem (including problem (2) that we consider) with any ambiguity set can be represented as one with an infinitely constrained ambiguity set. Apart from its generality, the infinitely constrained ambiguity set (4) is able to specify several interesting properties of probability distributions, including stochastic dominance, mean-dispersion, fourth moment, and entropic dominance—each of these has its own merit in characterizing the uncertainty. We refer interested readers to Chen et al. (2019) for more details on the modeling power of infinitely constrained ambiguity sets.

A key ingredient of solving problem (2) is the evaluation of the worst-case expected second-stage cost $\rho(\boldsymbol{x})$, for a fixed decision $\boldsymbol{x}$. Unfortunately, the possibly infinitely many expectation constraints render problem (3) becomes intractable, even when not accounting for adjustability (Chen et al. 2019). To tackle this issue of intractability, we consider a relaxed ambiguity set

$$\mathcal{F}_{\mathrm{R}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^I\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{q}, \tilde{\boldsymbol{z}})] \leq h(\boldsymbol{q}) \quad \forall \boldsymbol{q} \in \bar{\mathcal{Q}} \\ \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{W}] = 1 \end{array} \right\},$$

which involves a finite subset of expectation constraints parameterized by $\bar{\mathcal{Q}} = \{\boldsymbol{q}_j \in \mathcal{Q} : j \in [J]\}$. Based on the relaxed ambiguity set $\mathcal{F}_{\mathrm{R}}$, we also define a *lifted ambiguity set* $\mathcal{G}_{\mathrm{R}}$ that encompasses the primary random variable $\tilde{\boldsymbol{z}}$ and the auxiliary lifted random variable $\tilde{\boldsymbol{u}}$:

$$\mathcal{G}_{\mathrm{R}} = \left\{ \mathbb{Q} \in \mathcal{P}_0\left(\mathbb{R}^I \times \mathbb{R}^J\right) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \sim \mathbb{Q} & \\ \mathbb{E}_{\mathbb{Q}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu} & \\ \mathbb{E}_{\mathbb{Q}}[\tilde{u}_j] \leq h(\boldsymbol{q}_j) & \forall j \in [J] \\ \mathbb{Q}[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}] = 1 & \end{array} \right\},$$

where the *lifted support set* $\bar{\mathcal{W}}$ is defined as the epigraph of g together with the support set $\mathcal{W}$:

$$\bar{\mathcal{W}} = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^I \times \mathbb{R}^J \;\middle|\; \boldsymbol{z} \in \mathcal{W},\, g(\boldsymbol{q}_j, \boldsymbol{z}) \leq u_j \;\forall j \in [J] \right\}.$$

Throughout this paper we utilize the concept of conic representation and make the following assumption for tractability.

ASSUMPTION 1. *Given any finite $\bar{\mathcal{Q}}$, the conic representation of the set $\{(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} : \boldsymbol{z} = \boldsymbol{\mu}\}$ satisfies the Slater's condition.*

The lifted ambiguity set is first introduced by Wiesemann et al. (2014) for designing a standard form of the ambiguity set, where one of the key features is the neat expectation constraint that resides in an affine manifold. Indeed, the ambiguity sets $\mathcal{F}_{\mathrm{R}}$ and $\mathcal{G}_{\mathrm{R}}$ are equivalent in a way described in the following lemma proposed by Wiesemann et al. (2014).

LEMMA 2 (theorem 5, Wiesemann et al. 2014). *The ambiguity set $\mathcal{F}_{\mathrm{R}}$ is equivalent to the set of marginal distributions of $\tilde{\boldsymbol{z}}$ under all joint distribution $\mathbb{Q} \in \mathcal{G}_{\mathrm{R}}$, that is, $\mathcal{F}_{\mathrm{R}} = \cup_{\mathbb{Q}\in\mathcal{G}_{\mathrm{R}}}\{\Pi_{\tilde{\boldsymbol{z}}}\mathbb{Q}\}$.*

By virtue of Lemma 2, we have

$$\bar{\rho}(\boldsymbol{x}) = \sup_{\mathbb{Q}\in\mathcal{G}_{\mathrm{R}}} \mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] = \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{R}}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] \geq \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] = \rho(\boldsymbol{x}).$$

That is, concerning with an upper bound of the worst-case expected second-stage cost $\rho(\boldsymbol{x})$, ambiguity sets $\mathcal{F}_{\mathrm{R}}$ and $\mathcal{G}_{\mathrm{R}}$ are essentially the same. Quite notably, Bertsimas et al. (2019) show that inclusion of the auxiliary random variable $\tilde{\boldsymbol{u}}$ in $\mathcal{G}_{\mathrm{R}}$ would yield, in a systematic manner, an enhancement of the LDR approximation for the adjustable distributionally robust optimization problems. We will introduce this technique and adopt it to our setting in the next section.

3. (Extended) Linear Decision Rule Approximation

In the remainder of this paper, without loss of generality, we will focus on $\rho(\boldsymbol{x})$ in (3), which can be seamlessly incorporated into problem (2) for obtaining the optimal here-and-now decision. Under the condition of relatively complete and sufficiently expensive recourse, we can represent the objective function of the second-stage problem by exploring strong duality of a linear program: $f(\boldsymbol{x},\boldsymbol{z}) = \max_{k\in[K]}\left\{\boldsymbol{p}_k^\top(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x})\right\}$, where $\{\boldsymbol{p}_k\}_{k\in[K]}$ are extreme points of the polyhedron $\{\boldsymbol{p} \geq \boldsymbol{0} : \boldsymbol{B}^\top\boldsymbol{p} = \boldsymbol{d}\}$. The upper bound of $\rho(\boldsymbol{x})$ thus becomes

$$\bar{\rho}(\boldsymbol{x}) = \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{R}}} \mathbb{E}_{\mathbb{P}}\left[\max_{k\in[K]}\left\{\boldsymbol{p}_k^\top(\boldsymbol{b}(\tilde{\boldsymbol{z}}) - \boldsymbol{A}(\tilde{\boldsymbol{z}})\boldsymbol{x})\right\}\right], \tag{5}$$

which is the worst-case expectation of a convex and piecewise affine objective function over the relaxed ambiguity set $\mathcal{F}_{\mathrm{R}}$. Using the standard approach in distributionally robust optimization (see, *e.g.*, Delage and Ye 2010, Wiesemann et al. 2014), problem (5) can be reformulated as a conic program (Bertsimas et al. 2019). The resultant reformulation, however, is computationally expensive unless the number of extreme points is small. Hence, it is necessary and of practical interest to derive tractable approximations. To this end, first observe that we can also express problem (3) as a minimization problem over a measurable functional $\boldsymbol{y}$ as follows:

$$\bar{\rho}(\boldsymbol{x}) = \begin{cases} \min \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{R}}} \mathbb{E}_{\mathbb{P}}[\boldsymbol{d}^\top \boldsymbol{y}(\tilde{\boldsymbol{z}})] \\ \text{s.t. } \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{W} \\ \quad\ \boldsymbol{y} \in \mathcal{R}^{I,L}, \end{cases} \tag{6}$$

where $\mathcal{R}^{I,L}$ is the space of all measurable functions from $\mathbb{R}^I$ to $\mathbb{R}^L$. However, problem (6) is also computationally intractable because one is optimizing over arbitrary functions that reside in the infinite-dimensional space. Nevertheless, we can obtain an approximation from above by restricting $\boldsymbol{y}$ to a smaller class of functions. For instance, in the classical LDR approximation (Garstka and Wets 1974 and Ben-Tal et al. 2004), the admissible function is restricted to one that is affinely dependent on the primary random variable $\boldsymbol{z}$, *i.e.*, $\boldsymbol{y} \in \mathcal{L}^L$, where

$$\mathcal{L}^L = \left\{ \boldsymbol{y} \in \mathcal{R}^{I,L} \;\middle|\; \begin{array}{l} \exists \boldsymbol{y}_0,\, \boldsymbol{y}_{1i},\, \forall i \in [I]: \\ \boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{y}_0 + \sum_{i\in[I]} \boldsymbol{y}_{1i} z_i \end{array} \right\}.$$

Consequently, under the LDR approximation, we obtain an upper bound of $\bar{\rho}(\boldsymbol{x})$ by solving

$$\bar{\rho}_{\mathrm{LDR}}(\boldsymbol{x}) = \begin{cases} \min \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{R}}} \mathbb{E}_{\mathbb{P}}[\boldsymbol{d}^\top \boldsymbol{y}(\tilde{\boldsymbol{z}})] \\ \text{s.t. } \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{W} \\ \quad\ \boldsymbol{y} \in \mathcal{L}^L. \end{cases}$$

Bertsimas et al. (2019) recently introduce an enhancement of the LDR approximation, to which we refer as the extended linear decision rule (ELDR) approximation, by considering the lifted ambiguity set $\mathcal{G}_{\mathrm{R}}$ and an LDR with dependence on both $\tilde{\boldsymbol{z}}$ and the auxiliary random variable $\tilde{\boldsymbol{u}}$. Specially, with the ELDR approximation, one can solve

$$\bar{\rho}_{\mathrm{ELDR}}(\boldsymbol{x}) = \begin{cases} \min \sup_{\mathbb{Q}\in\mathcal{G}_{\mathrm{R}}} \mathbb{E}_{\mathbb{Q}}[\boldsymbol{d}^\top \boldsymbol{y}(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})] \\ \text{s.t. } \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}} \\ \quad\ \boldsymbol{y} \in \bar{\mathcal{L}}^L, \end{cases} \tag{7}$$

where the recourse decision $\boldsymbol{y}$ is restricted in the following class of affine functions:

$$\bar{\mathcal{L}}^L = \left\{ \boldsymbol{y} \in \mathcal{R}^{I+J,L} \;\middle|\; \begin{array}{l} \exists \boldsymbol{y}_0,\, \boldsymbol{y}_{1i},\, \boldsymbol{y}_{2j},\, \forall i \in [I],\, j \in [J]: \\ \boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) = \boldsymbol{y}_0 + \sum_{i\in[I]} \boldsymbol{y}_{1i} z_i + \sum_{j\in[J]} \boldsymbol{y}_{2j} u_j \end{array} \right\}.$$

As a larger class of functions, the ELDR approximation improves over the LDR approximation: that is, $\bar{\rho}_{\mathrm{LDR}}(\boldsymbol{x}) \geq \bar{\rho}_{\mathrm{ELDR}}(\boldsymbol{x}) \geq \bar{\rho}(\boldsymbol{x})$. Bertsimas et al. (2019) also point out that under the complete recourse assumption (and if necessary, some other mild conditions), by applying the ELDR approximation, the potential infeasibility issue of the LDR approximation can be resolved, and even the exact solution for adjustable problems with a one-dimensional adjustable decision (*i.e.*, $L=1$) can be obtained. We summarize the advantages of the ELDR approximation in the following result.

LEMMA 3. [theorems 2-4, Bertsimas et al. 2019] *The ELDR approximation of the worst-case second-stage cost* (6) *under* $\mathcal{G}_{\rm R}$*, problem* (7)*, possesses the following advantages:*

(i) It performs at least as well as the LDR approximation, i.e., $\bar{\rho}_{\rm LDR}(\boldsymbol{x}) \geq \bar{\rho}_{\rm ELDR}(\boldsymbol{x}) \geq \bar{\rho}(\boldsymbol{x})$.

(ii) If problem (1) *has complete and sufficiently expensive recourse, then for any ambiguity set* $\mathcal{F}_{\rm R}$ *such that* $\mathbb{E}_{\mathbb{P}}[|\tilde{z}_i|] < +\infty \ \forall i \in [I]$, $\mathbb{P} \in \mathcal{F}_{\rm R}$, *there exists a lifted ambiguity set* $\mathcal{G}_{\rm R}$ *whose corresponding ELDR approximation is feasible in problem* (7).

(iii) If problem (1) *has complete recourse and has only one second-stage decision variable, i.e.,* $L=1$*, then* $\rho(\boldsymbol{x}) = \bar{\rho}_{\rm ELDR}(\boldsymbol{x})$.

In the rest of this paper, we will work with the lifted ambiguity set $\mathcal{G}_{\rm R}$ and seek to improve the ELDR approximation $\bar{\rho}_{\rm ELDR}(\boldsymbol{x})$ as an upper bound of $\rho(\boldsymbol{x})$. Before proceeding, we conclude this section by showing that problem (7) can be reformulated as a tractable conic optimization problem, provided that problem (1) has relatively complete and sufficiently expensive recourse.

THEOREM 1. *Suppose the second-stage problem* (1) *has relatively complete and sufficiently expensive recourse. Then for any decision* $\boldsymbol{x} \in \mathcal{X}$*, problem* (7) *is equivalent to a conic program:*

$$\bar{\rho}_{\rm ELDR}(\boldsymbol{x}) = \left\{ \begin{array}{ll} \inf \ \boldsymbol{d}^\top \boldsymbol{y}_0 + \sum_{i\in[I]} \boldsymbol{d}^\top \boldsymbol{y}_{1i}\mu_i + \sum_{j\in[J]} \boldsymbol{d}^\top \boldsymbol{y}_{2j} h(\boldsymbol{q}_j) & \\ \text{s.t.} \ \boldsymbol{A}_0\boldsymbol{x} - \boldsymbol{b}_0 + \boldsymbol{B}\boldsymbol{y}_0 \geq \boldsymbol{t} & \\ \quad \left(\begin{bmatrix} \boldsymbol{A}_{1,m:}\boldsymbol{x} - b_{1m} + \boldsymbol{B}_{m:}\boldsymbol{y}_{11} \\ \vdots \\ \boldsymbol{A}_{I,m:}\boldsymbol{x} - b_{Im} + \boldsymbol{B}_{m:}\boldsymbol{y}_{1I} \end{bmatrix}, \begin{bmatrix} \boldsymbol{B}_{m:}\boldsymbol{y}_{21} \\ \vdots \\ \boldsymbol{B}_{m:}\boldsymbol{y}_{2J} \end{bmatrix}, t_m \right) \in \mathcal{K}^* & \forall m \in [M] \\ \quad \boldsymbol{t} \in \mathbb{R}^M, \ \boldsymbol{y}_0, \ \boldsymbol{y}_{1i}, \ \boldsymbol{y}_{2j} \in \mathbb{R}^L & \forall i \in [I], \ j \in [J], \end{array} \right. \tag{8}$$

where $\mathcal{K}^*$ *is the dual cone of* $\mathcal{K} = \mathbf{cl}\{(\boldsymbol{z},\boldsymbol{u},v) \in \mathbb{R}^I \times \mathbb{R}^J \times \mathbb{R} \mid (\boldsymbol{z},\boldsymbol{u})/v \in \bar{\mathcal{W}}, \ v > 0\}$*, and for* $i \in [I]$ *and* $m \in [M]$*,* $\boldsymbol{A}_{i,m:}$ *is the* m-th *row of the matrix* $\boldsymbol{A}_i$.

Proof of Theorem 1. Introducing dual variables α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$, the dual of problem (7) is

$$\bar{\rho}_{\rm ELDR}(\boldsymbol{x}) \leq \left\{ \begin{array}{ll} \inf \ \alpha + \boldsymbol{\beta}^\top \boldsymbol{\mu} + \sum_{j\in[J]} \gamma_j h(\boldsymbol{q}_j) & \\ \text{s.t.} \ \alpha + \boldsymbol{\beta}^\top \boldsymbol{z} + \boldsymbol{\gamma}^\top \boldsymbol{u} \geq \boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) & \forall (\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ \quad \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}) & \forall (\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ \quad \alpha \in \mathbb{R}, \ \boldsymbol{\beta} \in \mathbb{R}^I, \ \boldsymbol{\gamma} \in \mathbb{R}^J_+, \ \boldsymbol{y} \in \bar{\mathcal{L}}^L. & \end{array} \right. \tag{9}$$

Observe that each constraint in problem (9) requires a classical robust counterpart to be not less than a certain threshold. For instance, we can present the first constraint as

$$\boldsymbol{d}^\top \boldsymbol{y}_0 - \alpha \le \begin{cases} \inf & \boldsymbol{\beta}^\top \boldsymbol{z} + \boldsymbol{\gamma}^\top \boldsymbol{u} - \sum_{i \in [I]} \boldsymbol{d}^\top \boldsymbol{y}_{1i} z_i - \sum_{j \in [J]} \boldsymbol{d}^\top \boldsymbol{y}_{2j} u_j \\ \text{s.t.} & (\boldsymbol{z}, \boldsymbol{u}, 1) \succeq_{\mathcal{K}} \boldsymbol{0} \\ & \boldsymbol{z} \in \mathbb{R}^I,\ \boldsymbol{u} \in \mathbb{R}^J. \end{cases}$$

The dual of the right-hand side optimization problem is

$$\begin{aligned} \sup\ & -t_0 \\ \text{s.t.}\ & \left(\boldsymbol{\beta} - \begin{bmatrix} \boldsymbol{d}^\top \boldsymbol{y}_{11} \\ \vdots \\ \boldsymbol{d}^\top \boldsymbol{y}_{1I} \end{bmatrix}, \boldsymbol{\gamma} - \begin{bmatrix} \boldsymbol{d}^\top \boldsymbol{y}_{21} \\ \vdots \\ \boldsymbol{d}^\top \boldsymbol{y}_{2J} \end{bmatrix}, t_0 \right) \in \mathcal{K}^* \\ & t_0 \in \mathbb{R}, \end{aligned}$$

implying that the first constraint is satisfiable if the following constraint system is feasible:

$$\begin{cases} \alpha - \boldsymbol{d}^\top \boldsymbol{y}_0 - t_0 \ge 0 \\ \left(\boldsymbol{\beta} - \begin{bmatrix} \boldsymbol{d}^\top \boldsymbol{y}_{11} \\ \vdots \\ \boldsymbol{d}^\top \boldsymbol{y}_{1I} \end{bmatrix}, \boldsymbol{\gamma} - \begin{bmatrix} \boldsymbol{d}^\top \boldsymbol{y}_{21} \\ \vdots \\ \boldsymbol{d}^\top \boldsymbol{y}_{2J} \end{bmatrix}, t_0 \right) \in \mathcal{K}^*. \end{cases}$$

We can replace other constraints in a similar way and obtain:

$$\bar{\rho}_{\mathrm{ELDR}}(\boldsymbol{x}) \le \begin{cases} \inf\ \alpha + \boldsymbol{\beta}^\top \boldsymbol{\mu} + \sum_{j \in [J]} \gamma_j h(\boldsymbol{q}_j) \\ \text{s.t.}\ \alpha - \boldsymbol{d}^\top \boldsymbol{y}_0 - t_0 \ge 0 \\ \quad \left(\boldsymbol{\beta} - \begin{bmatrix} \boldsymbol{d}^\top \boldsymbol{y}_{11} \\ \vdots \\ \boldsymbol{d}^\top \boldsymbol{y}_{1I} \end{bmatrix}, \boldsymbol{\gamma} - \begin{bmatrix} \boldsymbol{d}^\top \boldsymbol{y}_{21} \\ \vdots \\ \boldsymbol{d}^\top \boldsymbol{y}_{2J} \end{bmatrix}, t_0 \right) \in \mathcal{K}^* \\ \quad \boldsymbol{A}_0 \boldsymbol{x} - \boldsymbol{b}_0 + \boldsymbol{B} \boldsymbol{y}_0 \ge \boldsymbol{t} \\ \quad \left(\begin{bmatrix} \boldsymbol{A}_{1,m:} \boldsymbol{x} - b_{1m} + \boldsymbol{B}_{m:} \boldsymbol{y}_{11} \\ \vdots \\ \boldsymbol{A}_{I,m:} \boldsymbol{x} - b_{Im} + \boldsymbol{B}_{m:} \boldsymbol{y}_{1I} \end{bmatrix}, \begin{bmatrix} \boldsymbol{B}_{m:} \boldsymbol{y}_{21} \\ \vdots \\ \boldsymbol{B}_{m:} \boldsymbol{y}_{2J} \end{bmatrix}, t_m \right) \in \mathcal{K}^* \qquad \forall m \in [M] \\ \quad \alpha \in \mathbb{R},\ \boldsymbol{\beta} \in \mathbb{R}^I,\ \boldsymbol{\gamma} \in \mathbb{R}^J_+,\ t_0 \in \mathbb{R},\ \boldsymbol{t} \in \mathbb{R}^M,\ \boldsymbol{y}_0,\ \boldsymbol{y}_{1i},\ \boldsymbol{y}_{2j} \in \mathbb{R}^L \quad \forall i \in [I],\ j \in [J]. \end{cases} \tag{10}$$

Since the optimal α of problem (10) takes a value of $\boldsymbol{d}^\top \boldsymbol{y}_0 + t_0$, we can, with some variable substitutions, plug the first two constraints into the objective and represent problem (10) as

$$
\begin{aligned}
\inf\ & \boldsymbol{d}^\top \boldsymbol{y}_0 + \sum_{i\in[I]} \boldsymbol{d}^\top \boldsymbol{y}_{1i}\mu_i + \sum_{j\in[J]} \boldsymbol{d}^\top \boldsymbol{y}_{2j} h(\boldsymbol{q}_j) + \left(t_0 + \boldsymbol{r}_0{}^\top \boldsymbol{\mu} + \sum_{j\in[J]} s_{0j} h(\boldsymbol{q}_j) \right) \\
\text{s.t.}\ & \boldsymbol{A}_0\boldsymbol{x} - \boldsymbol{b}_0 + \boldsymbol{B}\boldsymbol{y}_0 \geq \boldsymbol{t} \\
& \left(\begin{bmatrix} \boldsymbol{A}_{1,m:}\boldsymbol{x} - b_{1m} + \boldsymbol{B}_{m:}\boldsymbol{y}_{11} \\ \vdots \\ \boldsymbol{A}_{I,m:}\boldsymbol{x} - b_{Im} + \boldsymbol{B}_{m:}\boldsymbol{y}_{1I} \end{bmatrix}, \begin{bmatrix} \boldsymbol{B}_{m:}\boldsymbol{y}_{21} \\ \vdots \\ \boldsymbol{B}_{m:}\boldsymbol{y}_{2J} \end{bmatrix}, t_m \right) \in \mathcal{K}^* \qquad \forall m \in [M] \\
& \begin{bmatrix} \boldsymbol{d}^\top \boldsymbol{y}_{21} \\ \vdots \\ \boldsymbol{d}^\top \boldsymbol{y}_{2J} \end{bmatrix} + \boldsymbol{s}_0 \geq \boldsymbol{0} \\
& (\boldsymbol{r}_0, \boldsymbol{s}_0, t_0) \in \mathcal{K}^* \\
& \boldsymbol{r}_0 \in \mathbb{R}^I,\ \boldsymbol{s}_0 \in \mathbb{R}^J,\ t_0 \in \mathbb{R},\ \boldsymbol{t} \in \mathbb{R}^M,\ \boldsymbol{y}_0, \boldsymbol{y}_{1i}, \boldsymbol{y}_{2j} \in \mathbb{R}^L \qquad \forall i \in [I],\ j \in [J].
\end{aligned}
\tag{11}
$$

It remains to argue that it is free to remove the second last constraint and set $(\boldsymbol{r}_0, \boldsymbol{s}_0, t_0) = \boldsymbol{0}$ (that is, the last constraint can be removed, too). Observe that for any $(\boldsymbol{z}, \boldsymbol{u}, v) \in \mathcal{K}$, we have $(\boldsymbol{z}, \boldsymbol{u} + \boldsymbol{\delta}, v) \in \mathcal{K}\ \forall \boldsymbol{\delta} \geq \boldsymbol{0}$. By the definition of dual cone, for any $(\boldsymbol{r}, \boldsymbol{s}, t) \in \mathcal{K}^*$ and $\boldsymbol{\delta} \geq \boldsymbol{0}$, we have

$$
\begin{cases} \boldsymbol{r}^\top \boldsymbol{z} + \boldsymbol{s}^\top \boldsymbol{u} + tv \geq 0 \\ \boldsymbol{r}^\top \boldsymbol{z} + \boldsymbol{s}^\top \boldsymbol{u} + tv + \boldsymbol{s}^\top \boldsymbol{\delta} \geq 0, \end{cases}
$$

implying $\boldsymbol{s} \geq \boldsymbol{0}$. That is to say, $\boldsymbol{s} \geq \boldsymbol{0}$ holds for any $(\boldsymbol{r}, \boldsymbol{s}, t) \in \mathcal{K}^*$, which further implies $\boldsymbol{s}_0 \geq \boldsymbol{0}$ and $(\boldsymbol{B}_{m:}\boldsymbol{y}_{21}, \cdots, \boldsymbol{B}_{m:}\boldsymbol{y}_{2J}) \geq \boldsymbol{0}$ in (11). By Lemma 1, it then follows that $(\boldsymbol{d}^\top \boldsymbol{y}_{21}, \cdots, \boldsymbol{d}^\top \boldsymbol{y}_{2J}) \geq \boldsymbol{0}$. Hence, the second last constraint in (11) is redundant. The fourth term in the objective of (11) can be written as $(\boldsymbol{r}_0, \boldsymbol{s}_0, t_0)^\top (\boldsymbol{\mu}, h(\boldsymbol{q}_1), \cdots, h(\boldsymbol{q}_J), 1)$. Recall that $\boldsymbol{\mu} \in \mathcal{W}$ and $g(\boldsymbol{q}, \boldsymbol{\mu}) \leq h(\boldsymbol{q})$ for all $\boldsymbol{q} \in \mathcal{Q}$, we then have $(\boldsymbol{\mu}, h(\boldsymbol{q}_1), \cdots, h(\boldsymbol{q}_J), 1) \in \mathcal{K}$. Since (11) is a minimization problem and $(\boldsymbol{r}_0, \boldsymbol{s}_0, t_0) \in \mathcal{K}^*$, then $(\boldsymbol{r}_0, \boldsymbol{s}_0, t_0) = \boldsymbol{0}$ at optimality. We can now conclude that (11) is equivalent to the right-hand side of (8). Finally, throughout this proof, strong duality follows from the assumptions that the second-stage problem (1) has relatively complete and sufficiently expensive recourse, as well as the imposed Slater's condition; see details in (Bertsimas et al. 2019). □

4. Separation Problem for Tightening the Relaxation

To tighten the relaxation to the infinitely constrained ambiguity set, and ultimately, to improve the ELDR approximation as a tighter upper bound of the adjustable distributionally robust optimization problem, one natural idea is to include more expectation constraints in $\mathcal{F}_{\mathrm{R}}$. In this section, we propose a greedy procedure to effectively identify such expectation constraints.

Given an ambiguity set $\mathcal{G}_{\mathrm{R}}$, $\bar{\rho}_{\mathrm{ELDR}}(\boldsymbol{x})$ can be reformulated as problem (8), whose dual is given by the following conic optimization problem:[2]

$$
\bar{\rho}_{\mathrm{ELDR}}(x)=\left\{\begin{array}{lll}
\sup & \displaystyle\sum_{m\in[M]}\left\{\eta_m(b_{0m}-\boldsymbol{A}_{0,m:}\boldsymbol{x})+\sum_{i\in[I]}\xi_{mi}(b_{im}-\boldsymbol{A}_{i,m:}\boldsymbol{x})\right\} & \\
\text{s.t.} & \displaystyle\sum_{m\in[M]}\eta_m\boldsymbol{B}_{m:}^{\top}=\boldsymbol{d} & \\
& \displaystyle\sum_{m\in[M]}\xi_{mi}\boldsymbol{B}_{m:}^{\top}=\mu_i\boldsymbol{d} & \forall i\in[I]\\
& \displaystyle\sum_{m\in[M]}\zeta_{mj}\boldsymbol{B}_{m:}^{\top}=h(\boldsymbol{q}_j)\boldsymbol{d} & \forall j\in[J]\\
& (\boldsymbol{\xi}_m,\boldsymbol{\zeta}_m,\eta_m)\succeq_{\mathcal{K}}\boldsymbol{0} & \forall m\in[M]\\
& \boldsymbol{\xi}_m\in\mathbb{R}^I,\ \boldsymbol{\zeta}_m\in\mathbb{R}^J,\ \eta_m\in\mathbb{R} & \forall m\in[M],
\end{array}\right. \tag{12}
$$

where the set of conic constraints is equivalent to

$$
\frac{\boldsymbol{\xi}_m}{\eta_m}\in\mathcal{W}\quad\text{and}\quad g\left(\boldsymbol{q}_j,\frac{\boldsymbol{\xi}_m}{\eta_m}\right)\le\frac{\zeta_{mj}}{\eta_m}\quad\forall j\in[J],\ m\in[M].^3
$$

Each j-th expectation constraint $\mathbb{E}_{\mathbb{P}}[g(\boldsymbol{q}_j,\tilde{\boldsymbol{z}})]\le h(\boldsymbol{q}_j)$ in the relaxed ambiguity set $\mathcal{F}_{\mathrm{R}}$ leads to, in problem (12), a collection of decision variables $\{\xi_{mj}\}_{m\in[M]}$ as well as two collections of constraints $\sum_{m\in[M]}\zeta_{mj}\boldsymbol{B}_{m:}^{\top}=h(\boldsymbol{q}_j)$ and $g(\boldsymbol{q}_j,\boldsymbol{\xi}_m/\eta_m)\le\zeta_{mj}/\eta_m$. Inspired by this observation, given the optimal solution $(\eta_m^\star,\boldsymbol{\xi}_m^\star)_{m\in[M]}$ to problem (12), we can identify a violated expectation constraint $\mathbb{E}_{\mathbb{P}}[g(\boldsymbol{q},\tilde{\boldsymbol{z}})]>h(\boldsymbol{q})$ for some $\boldsymbol{q}\in\mathcal{Q}\setminus\bar{\mathcal{Q}}$, if the following system is infeasible:

$$
\left\{\begin{array}{l}
\displaystyle\sum_{m\in[M]}\zeta_m\boldsymbol{B}_{m:}^{\top}=h(\boldsymbol{q})\boldsymbol{d}\\
\eta_m^\star\cdot g\left(\boldsymbol{q},\dfrac{\boldsymbol{\xi}_m^\star}{\eta_m^\star}\right)\le\zeta_m\quad\forall m\in[M];
\end{array}\right.
$$

or equivalently, if the following linear program over $\boldsymbol{\zeta}$, given the particular parameter $\boldsymbol{q}$, is infeasible:

$$
\begin{array}{lll}
\min & 0 & \\
\text{s.t.} & \displaystyle\sum_{m\in[M]}\zeta_m\boldsymbol{B}_{m:}^{\top}=h(\boldsymbol{q})\boldsymbol{d} & \\
& \zeta_m\ge\theta_m(\boldsymbol{q}) & \forall m\in[M]\\
& \boldsymbol{\zeta}\in\mathbb{R}^M, &
\end{array} \tag{13}
$$

where given $(\eta_m^\star,\boldsymbol{\xi}_m^\star)$, we denote $\theta_m(\boldsymbol{q})=\eta_m^\star\cdot g(\boldsymbol{q},\boldsymbol{\xi}_m^\star/\eta_m^\star)$ for each $m\in[M]$.

We refer to the following dual of problem (13) as the indicator problem:

$$
\max_{\boldsymbol{\lambda}\in\mathrm{recc}(\boldsymbol{B})}\sum_{m\in[M]}\theta_m(\boldsymbol{q})\boldsymbol{B}_{m:}\boldsymbol{\lambda}-h(\boldsymbol{q})\boldsymbol{d}^{\top}\boldsymbol{\lambda}, \tag{14}
$$

[2] Strong duality herein is a byproduct of the established strong duality in Theorem 1; see, *e.g.*, Bertsimas et al. (2019).

[3] It is not hard to argue that $\boldsymbol{\xi}_m^\star=\boldsymbol{0}$ whenever $\eta_m^\star=0$, because otherwise the boundedness of $\mathcal{W}$ would be violated.

where $\mathrm{recc}(\boldsymbol{B}) = \{\boldsymbol{\lambda} \in \mathbb{R}^L \mid \boldsymbol{B}\boldsymbol{\lambda} \geq \mathbf{0}\}$ is the recession cone generated by the recourse matrix $\boldsymbol{B}$. Since problem (14) is always feasible, its objective goes to positive infinity whenever problem (13) is infeasible, indicating a violated expectation constraint $\mathbb{E}_{\mathbb{P}}[g(\boldsymbol{q}, \tilde{\boldsymbol{z}})] > h(\boldsymbol{q})$. In addition, observe that for a particular $\boldsymbol{q} \in \mathcal{Q} \setminus \bar{\mathcal{Q}}$, the corresponding indicator problem (14) is unbounded if and only if some extreme ray $\boldsymbol{\lambda}^\star$ of $\mathrm{recc}(\boldsymbol{B})$ satisfies $\sum_{m\in[M]} \theta_m(\boldsymbol{q})\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star - h(\boldsymbol{q})\boldsymbol{d}^\top\boldsymbol{\lambda}^\star > 0$. In summary, to find a $\boldsymbol{q} \in \mathcal{Q} \setminus \bar{\mathcal{Q}}$ that leads to a violated expectation constraint to be included in the relaxed ambiguity set, it is equivalent to verify whether for *some* extreme ray $\boldsymbol{\lambda}^\star$ of $\mathrm{recc}(\boldsymbol{B})$,

$$\max_{\boldsymbol{q}\in\mathcal{Q}} \sum_{m\in[M]} \theta_m(\boldsymbol{q})\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star - h(\boldsymbol{q})\boldsymbol{d}^\top\boldsymbol{\lambda}^\star > 0 \quad \Longleftrightarrow \quad \min_{\boldsymbol{q}\in\mathcal{Q}} h(\boldsymbol{q})\boldsymbol{d}^\top\boldsymbol{\lambda}^\star - \sum_{m\in[M]} \theta_m(\boldsymbol{q})\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star < 0.$$

Therefore, to identify a violated expectation constraint, it is sufficient to focus on the finitely many extreme rays of $\mathrm{recc}(\boldsymbol{B})$ and for each extreme ray $\boldsymbol{\lambda}^\star$, solve the separation problem

$$\max_{\boldsymbol{q}\in\mathcal{Q}} \sum_{m\in[M]} \theta_m(\boldsymbol{q})\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star - h(\boldsymbol{q})\boldsymbol{d}^\top\boldsymbol{\lambda}^\star \quad \Longleftrightarrow \quad \min_{\boldsymbol{q}\in\mathcal{Q}} h(\boldsymbol{q})\boldsymbol{d}^\top\boldsymbol{\lambda}^\star - \sum_{m\in[M]} \theta_m(\boldsymbol{q})\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star. \tag{15}$$

Note that given $\mathrm{recc}(\boldsymbol{B})$, the number of its extreme rays is known to be always finite and they can be easily identified in many specific problems such as our three experiments in Section 6; see, for example, Definition 4 and Theorem 4. In correspondence to an extreme ray $\boldsymbol{\lambda}^\star$ of $\mathrm{recc}(\boldsymbol{B})$, there exists $\mathbb{Q}^\star \in \mathcal{G}_{\mathrm{R}}$ such that the objective function of the maximization problem in (15) can be read as $\boldsymbol{d}^\top\boldsymbol{\lambda}^\star(\mathbb{E}_{\mathbb{Q}^\star}[g(\boldsymbol{q}, \tilde{\boldsymbol{z}})] - h(\boldsymbol{q}))$. As a result, the separation problem (15) becomes

$$\max_{\boldsymbol{q}\in\mathcal{Q}} \boldsymbol{d}^\top\boldsymbol{\lambda}^\star(\mathbb{E}_{\mathbb{Q}^\star}[g(\boldsymbol{q}, \tilde{\boldsymbol{z}})] - h(\boldsymbol{q})) \quad \Longleftrightarrow \quad \min_{\boldsymbol{q}\in\mathcal{Q}} \boldsymbol{d}^\top\boldsymbol{\lambda}^\star(h(\boldsymbol{q}) - \mathbb{E}_{\mathbb{Q}^\star}[g(\boldsymbol{q}, \tilde{\boldsymbol{z}})]).$$

Indeed, such a distribution $\mathbb{Q}^\star$ can be interpreted as the worst-case distribution in the ambiguity set $\mathcal{G}_{\mathrm{R}}$ because the value of $\bar{\rho}_{\mathrm{ELDR}}(\boldsymbol{x})$ coincides with the expectation of the optimal ELDR approximation with respect to this distribution $\mathbb{Q}^\star$. We formalize these results as follows.

THEOREM 2. *Suppose the second-stage problem* (1) *has relatively complete and sufficiently expensive recourse. Let* $(\boldsymbol{\xi}_m^\star, \boldsymbol{\zeta}_m^\star, \eta_m^\star)_{m\in[M]}$ *be the optimal solution of problem* (12) *and* $\boldsymbol{\lambda}^\star$ *be an extreme ray of* $\mathrm{recc}(\boldsymbol{B})$. *We have:*

(i) The probability distribution defined as

$$\mathbb{Q}^\star\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) = \left(\frac{\boldsymbol{\xi}_m^\star}{\eta_m^\star}, \frac{\boldsymbol{\zeta}_m^\star}{\eta_m^\star}\right)\right] = \frac{\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star}\eta_m^\star \quad \forall m \in [M] : \eta_m^\star > 0$$

resides in $\mathcal{G}_{\mathrm{R}}$*, that is,* $\mathbb{Q}^\star \in \mathcal{G}_{\mathrm{R}}$*. The expected second-stage cost under* $\mathbb{Q}^\star$ *is bounded from above by* $\bar{\rho}_{\mathrm{ELDR}}(\boldsymbol{x})$*, that is,* $\mathbb{E}_{\mathbb{Q}^\star}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] \leq \bar{\rho}_{\mathrm{ELDR}}(\boldsymbol{x})$.

(ii) The objective function of problem (15) *in the left-hand side presentation as a maximization problem can be represented as* $\boldsymbol{d}^\top\boldsymbol{\lambda}^\star(\mathbb{E}_{\mathbb{Q}^\star}[g(\boldsymbol{q}, \tilde{\boldsymbol{z}})] - h(\boldsymbol{q}))$.

(iii) Let $\boldsymbol{y}^\star$ *be the optimal ELDR approximation obtained from solving problem* (8)*, then* $\mathbb{E}_{\mathbb{Q}^\star}[\boldsymbol{d}^\top\boldsymbol{y}^\star(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{u}})] = \boldsymbol{d}^\top\boldsymbol{y}_0^\star + \sum_{i\in[I]}\boldsymbol{d}^\top\boldsymbol{y}_{1i}^\star\mu_i + \sum_{j\in[J]}\boldsymbol{d}^\top\boldsymbol{y}_{2j}^\star h(\boldsymbol{q}_j) = \bar{\rho}_{\mathrm{ELDR}}(\boldsymbol{x})$.

Proof of Theorem 2. We only need to focus on the case where every extreme ray $\boldsymbol{\lambda}^\star$ satisfies $\boldsymbol{d}^\top\boldsymbol{\lambda}^\star > 0$. If $\boldsymbol{d}^\top\boldsymbol{\lambda}^\star = 0$, then $\boldsymbol{B}\boldsymbol{\lambda}^\star \geq \boldsymbol{0}$ implies $\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star = 0\ \forall m\in[M]$, where we recall from Lemma 1 that $\boldsymbol{d}$ is a non-negative linear combination of rows of $\boldsymbol{B}$ under the sufficiently expensive recourse condition. Such a case is trivial since the objective of problem (15) is always zero for any $\boldsymbol{q}\in\mathcal{Q}$.

In view of (i), we directly verify $\mathbb{Q}^\star\in\mathcal{G}_{\mathrm{R}}$. Firstly, the support constraint naturally follows. Secondly, for the probability constraint, we observe that

$$\sum_{m\in[M]:\eta_m^\star>0}\frac{\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star}\eta_m^\star = \frac{\sum_{m\in[M]}\eta_m^\star\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star} = \frac{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star} = 1.$$

Thirdly, for the expectation constraint, we have

$$\mathbb{E}_{\mathbb{Q}^\star}[\tilde{\boldsymbol{z}}] = \sum_{m\in[M]:\eta_m^\star>0}\frac{\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star}\eta_m^\star\frac{\boldsymbol{\xi}_m^\star}{\eta_m^\star} = \frac{\sum_{m\in[M]}\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star\boldsymbol{\xi}_m^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star} = \boldsymbol{\mu},$$

and for each j-th expectation constraint, we have

$$\mathbb{E}_{\mathbb{Q}^\star}[g(\boldsymbol{q}_j,\tilde{\boldsymbol{z}})] = \sum_{m\in[M]:\eta_m^\star>0}\frac{\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star}\eta_m^\star g\left(\boldsymbol{q}_j,\frac{\boldsymbol{\xi}_m^\star}{\eta_m^\star}\right) \leq \sum_{m\in[M]:\eta_m^\star>0}\frac{\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star}\eta_m^\star\frac{\zeta_{mj}^\star}{\eta_m^\star} = h(\boldsymbol{q}_j).$$

Finally, $\mathbb{Q}^\star\in\mathcal{G}_{\mathrm{R}}$ yields $\mathbb{E}_{\mathbb{Q}^\star}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] \leq \sup_{\mathbb{Q}\in\mathcal{G}_{\mathrm{R}}}\mathbb{E}_{\mathbb{Q}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] = \bar{\rho}(\boldsymbol{x}) \leq \bar{\rho}_{\mathrm{ELDR}}(\boldsymbol{x})$.

As for (ii), it is sufficient to note that

$$\begin{aligned}\textstyle\sum_{m\in[M]}\theta_m(\boldsymbol{q})\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star &= \boldsymbol{d}^\top\boldsymbol{\lambda}^\star\textstyle\sum_{m\in[M]}\dfrac{\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star}\eta_m^\star g\left(\boldsymbol{q},\dfrac{\boldsymbol{\xi}_m^\star}{\eta_m^\star}\right)\\ &= \boldsymbol{d}^\top\boldsymbol{\lambda}^\star\textstyle\sum_{m\in[M]:\eta_m^\star>0}\dfrac{\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star}\eta_m^\star g\left(\boldsymbol{q},\dfrac{\boldsymbol{\xi}_m^\star}{\eta_m^\star}\right)\\ &= \boldsymbol{d}^\top\boldsymbol{\lambda}^\star\mathbb{E}_{\mathbb{Q}^\star}[g(\boldsymbol{q},\tilde{\boldsymbol{z}})].\end{aligned}$$

In view of (iii), we compute the value of $\mathbb{E}_{\mathbb{Q}^\star}[\boldsymbol{d}^\top\boldsymbol{y}^\star(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{u}})]$ as follows:

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}^\star}[\boldsymbol{d}^\top\boldsymbol{y}^\star(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{u}})] &= \sum_{m\in[M]:\eta_m^\star>0}\frac{\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star}\eta_m^\star\left(\boldsymbol{d}^\top\boldsymbol{y}_0^\star + \sum_{i\in[I]}\boldsymbol{d}^\top\boldsymbol{y}_{1i}^\star\frac{\xi_{mi}^\star}{\eta_m^\star} + \sum_{j\in[J]}\boldsymbol{d}^\top\boldsymbol{y}_{2j}^\star\frac{\zeta_{mj}^\star}{\eta_m^\star}\right)\\ &= \boldsymbol{d}^\top\boldsymbol{y}_0^\star + \sum_{m\in[M]:\eta_m^\star>0}\frac{\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star}\eta_m^\star\left(\sum_{i\in[I]}\boldsymbol{d}^\top\boldsymbol{y}_{1i}^\star\frac{\xi_{mi}^\star}{\eta_m^\star} + \sum_{j\in[J]}\boldsymbol{d}^\top\boldsymbol{y}_{2j}^\star\frac{\zeta_{mj}^\star}{\eta_m^\star}\right)\\ &= \boldsymbol{d}^\top\boldsymbol{y}_0^\star + \sum_{i\in[I]}\boldsymbol{d}^\top\boldsymbol{y}_{1i}^\star\left(\sum_{m\in[M]}\frac{\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star}\xi_{mi}^\star\right) + \sum_{j\in[J]}\boldsymbol{d}^\top\boldsymbol{y}_{2j}^\star\left(\sum_{m\in[M]}\frac{\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star}{\boldsymbol{d}^\top\boldsymbol{\lambda}^\star}\zeta_{mj}^\star\right)\\ &= \boldsymbol{d}^\top\boldsymbol{y}_0^\star + \sum_{i\in[I]}\boldsymbol{d}^\top\boldsymbol{y}_{1i}^\star\mu_i + \sum_{j\in[J]}\boldsymbol{d}^\top\boldsymbol{y}_{2j}^\star h(\boldsymbol{q}_j),\end{aligned}$$

which coincides with $\bar{\rho}_{\mathrm{ELDR}}(\boldsymbol{x})$. □

Algorithm 1: GIP for Tightening the ELDR Approximation for Problem (2)

Input: An initial finite subset $\bar{\mathcal{Q}} \subseteq \mathcal{Q}$ and a maximal number of iterations

repeat

  step 1: Solve problem (7) with $\mathcal{G}_{\mathrm{R}}$ parameterized by $\bar{\mathcal{Q}}$ and obtain solution $\boldsymbol{x}^\star$.

  step 2: Given the optimal solution $\boldsymbol{x}^\star$, solve problem (12) to obtain $(\eta_m^\star, \boldsymbol{\xi}_m^\star)_{m\in[M]}$.

  step 3: Compute $\boldsymbol{q}$ such that the corresponding indicator problem (14) is unbounded and update $\bar{\mathcal{Q}} = \bar{\mathcal{Q}} \cup \{\boldsymbol{q}\}$ in $\mathcal{G}_{\mathrm{R}}$.

until problem (14) is bounded for all $\boldsymbol{q} \in \mathcal{Q}$ or the maximal number of iterations is reached;

Output: Solution $\boldsymbol{x}^\star$

To conclude this section, we propose an algorithm, called greedy improvement procedure (GIP), which leverages the unboundedness of the indicator problem for tightening the ELDR upper bound of the desired $\rho(\boldsymbol{x})$. Algorithm 1 summarizes the procedure of how GIP solves problem (2), where the idea is to tighten the ambiguity set (towards the infinitely constrained one) while at the same time, enhancing the ELDR approximation. Specifically, on the one hand, we identify violated expectation constraints to be included in the relaxed ambiguity set $\mathcal{F}_{\mathrm{R}}$ (steps 2 and 3), so as to obtain a tighter relaxation of the infinitely constrained ambiguity set $\mathcal{F}$; on the other hand, additional auxiliary random variables in the lifted ambiguity set that corresponds to newly added constraints contribute to gradually improving the approximation quality of ELDR (step 1).

We remark that in correspondence to various choices of the infinitely constrained ambiguity set, computing $\boldsymbol{q}$ in step 3 may reduce to different types of problems. For example, when using a covariance dominance ambiguity set introduced subsequently in Section 6.1, for each extreme ray $\boldsymbol{\lambda}^\star$ of $\mathrm{recc}(\boldsymbol{B})$ (and the respective worst-case distribution $\mathbb{Q}^\star$ as given in Theorem 2), the separation problem (15) reduces to

$$\min_{\|\boldsymbol{q}\|_2 \leq 1} \boldsymbol{d}^\top \boldsymbol{\lambda}^\star (\boldsymbol{q}^\top \boldsymbol{\Sigma} \boldsymbol{q} - \mathbb{E}_{\mathbb{Q}^\star}[(\boldsymbol{q}^\top(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))^2]) \quad \Longleftrightarrow \quad \min_{\boldsymbol{q}} \boldsymbol{q}^\top (\boldsymbol{\Sigma} - \mathbb{E}_{\mathbb{Q}^\star}[(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})(\tilde{\boldsymbol{z}} - \boldsymbol{\mu})^\top])\boldsymbol{q},$$

which is a minimal eigenvalue problem that can be solved efficiently. For another example, when considering a fourth moment ambiguity set in the coming Section 6.3, the separation problem reduces to finding the minimum of a multivariate polynomial—a problem that can be solved by using numerical optimization techniques such as the trust region method.

5. Generalization to Multi-Stage Problems

Although two-stage and multi-stage adjustable problems appear to be similar, their computational complexities differs considerably (Georghiou et al. 2015). For instance, obtaining the exact solutions of two-stage linear problems is already #P-hard, while multi-stage adjustable problems are believed to be *"computationally intractable already when medium-accuracy solutions are sought"* (Shapiro and Nemirovski 2005). A key challenge of solving multi-stage problems is to include non-anticipativity constraints that are necessary to capture the nature of multi-stage decisions, which should be (only) adjustable to sequentially revealed information. Fortunately, our approach to iteratively improves the ELDR approximation can be extended to multi-stage problems.

Let us consider a $(T+1)$-stage problem. For every $t\in[T]$, in processing from stage t to stage $(t+1)$, the uncertain components z_i with $i\in[I_t]\setminus[I_{t-1}]$ of the overall uncertainty $\boldsymbol{z}$ reveal, where $0=I_0<I_1<\cdots<I_T=I$. We consider the following infinitely constrained ambiguity set

$$\mathcal{F}=\left\{\mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^I\right)\;\middle|\;\begin{array}{l}\tilde{\boldsymbol{z}}\sim\mathbb{P}\\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}]=\boldsymbol{\mu}\\ \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{q},\tilde{\boldsymbol{z}})]\leq h(\boldsymbol{q})\quad\forall\boldsymbol{q}\in\mathcal{Q}_t,\,t\in[T]\\ \mathbb{P}[\tilde{\boldsymbol{z}}\in\mathcal{W}]=1\end{array}\right\}.$$

For each $t\in[T]$, let $\bar{\mathcal{Q}}_t=\{\boldsymbol{q}_{tj}\}_{j\in[J_t]}\subseteq\mathcal{Q}_t$. Then we have the relaxed ambiguity set

$$\mathcal{F}_{\mathrm{R}}=\left\{\mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^I\right)\;\middle|\;\begin{array}{l}\tilde{\boldsymbol{z}}\sim\mathbb{P}\\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}]=\boldsymbol{\mu}\\ \mathbb{E}_{\mathbb{P}}[g(\boldsymbol{q}_{tj},\tilde{\boldsymbol{z}})]\leq h(\boldsymbol{q}_{tj})\quad\forall t\in[T],\,j\in[J_t]\\ \mathbb{P}[\tilde{\boldsymbol{z}}\in\mathcal{W}]=1\end{array}\right\}$$

and the corresponding lifted ambiguity set $\mathcal{G}_{\mathrm{R}}$ that encompasses the auxiliary random variable:

$$\mathcal{G}_{\mathrm{R}}=\left\{\mathbb{Q}\in\mathcal{P}_0\left(\mathbb{R}^I\times\mathbb{R}^J\right)\;\middle|\;\begin{array}{ll}(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{u}})\sim\mathbb{Q}&\\ \mathbb{E}_{\mathbb{Q}}[\tilde{\boldsymbol{z}}]=\boldsymbol{\mu}&\\ \mathbb{E}_{\mathbb{Q}}[\tilde{u}_{tj}]\leq h(\boldsymbol{q}_{tj})&\forall t\in[T],\,j\in[J_t]\\ \mathbb{Q}\left[(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{u}})\in\bar{\mathcal{W}}\right]=1&\end{array}\right\},$$

where $J=\sum_{t\in[T]}J_t$ and $\bar{\mathcal{W}}=\{(\boldsymbol{z},\boldsymbol{u})\in\mathbb{R}^I\times\mathbb{R}^J\mid\boldsymbol{z}\in\mathcal{W},\;g(\boldsymbol{q}_{tj},\boldsymbol{z})\leq u_{tj}\;\forall t\in[T],\,j\in[J_t]\}$.

Given the subsets $\mathcal{I}_\ell\subseteq[I]$ that reflect the information dependency of the adjustable decisions y_ℓ for every $\ell\in[L]$, we consider the generalization of problem (6) as follows:

$$\bar{\phi}(\boldsymbol{x})=\left\{\begin{array}{lll}\min&\sup\limits_{\mathbb{P}\in\mathcal{F}_{\mathrm{R}}}\mathbb{E}_{\mathbb{P}}[\boldsymbol{d}^\top\boldsymbol{y}(\tilde{\boldsymbol{z}})]&\\ \text{s.t.}&\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}(\boldsymbol{z})\geq\boldsymbol{b}(\boldsymbol{z})&\forall\boldsymbol{z}\in\mathcal{W}\\ &y_\ell\in\mathcal{R}^I(\mathcal{I}_\ell)&\forall\ell\in[L],\end{array}\right.\tag{16}$$

where we define the space of restricted measurable functions as

$$\mathcal{R}^I(\mathcal{I})=\left\{y\in\mathcal{R}^{I,1}\;\middle|\;y(\boldsymbol{v})=y(\boldsymbol{w})\;\forall\boldsymbol{v},\boldsymbol{w}\in\mathbb{R}^I:v_k=w_k\;\forall k\in\mathcal{I}\right\}.$$

Problem (16) solves for the optimal decision rule that takes into account non-anticipativity. Note that without loss of generality, we can assume $\emptyset \neq \mathcal{I}_1 \subseteq \mathcal{I}_2 \subseteq \cdots \subseteq \mathcal{I}_L = [I]$. Following the spirit of ELDR approximation, we consider an upper bound of $\bar{\phi}(\boldsymbol{x})$ as follows:

$$\bar{\phi}_{\mathrm{ELDR}}(\boldsymbol{x}) = \begin{cases} \min \sup\limits_{\mathbb{P} \in \mathcal{G}_{\mathrm{R}}} \mathbb{E}_{\mathbb{P}}[\boldsymbol{d}^\top \boldsymbol{y}(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})] \\ \text{s.t. } \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}) & \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}} \\ \quad y_\ell \in \bar{\mathcal{L}}(\mathcal{I}_\ell, \mathcal{J}_\ell) & \forall \ell \in [L], \end{cases}$$

where the LDR takes the form of

$$\bar{\mathcal{L}}(\mathcal{I}, \mathcal{J}) = \left\{ y \in \mathcal{R}^{I+J,1} \;\middle|\; \begin{array}{l} \exists y_0,\, y_{1i},\, y_{2j} \in \mathbb{R},\, \forall i \in \mathcal{I},\, j \in \mathcal{J}: \\ y(\boldsymbol{z}, \boldsymbol{u}) = y_0 + \sum_{i \in \mathcal{I}} y_{1i} z_i + \sum_{j \in \mathcal{J}} y_{2j} u_j \end{array} \right\}$$

and subsets $\mathcal{J}_\ell \subseteq [J]$, $\ell \in [L]$ are consistent with the information restriction imposed by $\mathcal{I}_\ell \subseteq [I]$. Applying similar techniques as in the proof of Theorem 1, $\bar{\phi}_{\mathrm{ELDR}}(\boldsymbol{x})$ is equivalent to a conic program:

$$\bar{\phi}_{\mathrm{ELDR}}(\boldsymbol{x}) = \begin{cases} \inf\ \boldsymbol{d}^\top \boldsymbol{y}_0 + \sum\limits_{i \in [I]} \boldsymbol{d}^\top \boldsymbol{y}_{1i} \mu_i + \sum\limits_{j \in [J]} \boldsymbol{d}^\top \boldsymbol{y}_{2j} h(\boldsymbol{q}_j) \\ \text{s.t. } \boldsymbol{A}_{0,m:}\boldsymbol{x} - b_{0m} + \boldsymbol{B}_{m:}\boldsymbol{y}_0 - t_m \geq 0 & \forall m \in [M] \\ \quad \begin{bmatrix} \boldsymbol{A}_{1,m:}\boldsymbol{x} - b_{1m} + \boldsymbol{B}_{m:}\boldsymbol{y}_{11} \\ \vdots \\ \boldsymbol{A}_{I,m:}\boldsymbol{x} - b_{Im} + \boldsymbol{B}_{m:}\boldsymbol{y}_{1I} \end{bmatrix} - \boldsymbol{r}_m = \boldsymbol{0} & \forall m \in [M] \\ \quad \begin{bmatrix} \boldsymbol{B}_{m:}\boldsymbol{y}_{21} \\ \vdots \\ \boldsymbol{B}_{m:}\boldsymbol{y}_{2J} \end{bmatrix} - \boldsymbol{s}_m = \boldsymbol{0} & \forall m \in [M] \\ \quad (\boldsymbol{r}_m, \boldsymbol{s}_m, t_m) \succeq_{\mathcal{K}^*} \boldsymbol{0} & \forall m \in [M] \\ \quad y_{1i\ell} = 0 & \forall \ell \in [L],\, i \in [I] \setminus \mathcal{I}_\ell \\ \quad y_{2j\ell} = 0 & \forall \ell \in [L],\, j \in [J] \setminus \mathcal{J}_\ell \\ \quad \boldsymbol{r}_m \in \mathbb{R}^I,\, \boldsymbol{s}_m \in \mathbb{R}^J,\, \boldsymbol{t} \in \mathbb{R}^M,\, \boldsymbol{y}_0, \boldsymbol{y}_{1i}, \boldsymbol{y}_{2j} \in \mathbb{R}^L & \forall i \in [I],\, j \in [J],\, m \in [M], \end{cases}$$

provided that the problem has relatively complete and sufficiently expensive recourse.

Taking dual of $\bar{\phi}_{\text{ELDR}}(\boldsymbol{x})$, we obtain

$$\bar{\phi}_{\text{ELDR}}(\boldsymbol{x}) = \begin{cases} \sup \sum_{m\in[M]} \left\{ \eta_m (b_{0m} - \boldsymbol{A}_{0,m:}\boldsymbol{x}) + \sum_{i\in[I]} \xi_{mi}(b_{im} - \boldsymbol{A}_{i,m:}\boldsymbol{x}) \right\} \\ \text{s.t.} \sum_{m\in[M]} \eta_m \boldsymbol{B}_{m:}^\top = \boldsymbol{d} \\ \quad \sum_{m\in[M]} \xi_{mi} B_{m\ell} = \mu_i d_\ell & \forall \ell\in[L],\ i\in\mathcal{I}_\ell \\ \quad \sum_{m\in[M]} \zeta_{mj} B_{m\ell} = h(\boldsymbol{q}_j) d_\ell & \forall \ell\in[L],\ j\in\mathcal{J}_\ell \\ \quad (\boldsymbol{\xi}_m, \boldsymbol{\zeta}_m, \eta_m) \succeq_{\mathcal{K}} \boldsymbol{0} & \forall m\in[M] \\ \quad \boldsymbol{\xi}_m\in\mathbb{R}^I,\ \boldsymbol{\zeta}_m\in\mathbb{R}^J,\ \eta_m\in\mathbb{R} & \forall m\in[M]. \end{cases}$$

Given the optimal solution $(\eta_m^\star, \boldsymbol{\xi}_m^\star)_{m\in[M]}$, we can identify a violated expectation constraint if the following optimization problem is infeasible for a particular $\boldsymbol{q}^\star\in\mathcal{Q}_t$ for some $t\in[T]$:

$$\begin{array}{lll} \min 0 \\ \text{s.t.} \sum_{m\in[M]} \zeta_m B_{m\ell} = h(\boldsymbol{q}^\star) d_\ell & \forall \ell\in[L]: \kappa_\ell(\boldsymbol{q}^\star)=1 \\ \quad \zeta_m \ge \theta_m(\boldsymbol{q}^\star) & \forall m\in[M] \\ \quad \boldsymbol{\zeta}\in\mathbb{R}^M, \end{array} \tag{17}$$

where for all $\ell\in[L]$, $\kappa_\ell:\mathcal{Q}\mapsto\{0,1\}$ is an indicator function such that $\kappa_\ell(\boldsymbol{q}^\star)=1$ means that the adjustable decision y_ℓ depends on $\tilde{u}^\star$ associated with the expectation constraint $\mathbb{E}_{\mathbb{P}}[g(\boldsymbol{q}^\star,\tilde{\boldsymbol{z}})]\le h(\boldsymbol{q}^\star)$. Given $(\eta_m^\star,\boldsymbol{\xi}_m^\star)$, we denote $\theta_m(\boldsymbol{q}) = \eta_m^\star\cdot g(\boldsymbol{q},\boldsymbol{\xi}_m^\star/\eta_m^\star)$ for all $m\in[M]$. Following from aforementioned set-ups, we assume that $\boldsymbol{q}^\star\in\mathcal{Q}_t$ implies

$$\kappa_\ell(\boldsymbol{q}^\star) = \begin{cases} 1 & [I_t]\subseteq\mathcal{I}_\ell \\ 0 & \text{otherwise.} \end{cases}$$

Let $L^\star$ be the cardinality of the set $\{\ell\in[L]:\kappa_\ell(\boldsymbol{q}^\star)=1\}$, $\boldsymbol{B}^\star\in\mathbb{R}^{M,L^\star}$ be the sub-matrix of $\boldsymbol{B}$ whose columns correspond to those non-zero columns in $\boldsymbol{B}\text{diag}(\kappa_1(\boldsymbol{q}^\star),\kappa_2(\boldsymbol{q}^\star),\ldots,\kappa_L(\boldsymbol{q}^\star))$, and $\boldsymbol{d}^\star$ be the sub-vector of $\boldsymbol{d}$ whose elements correspond to those non-zero components in $\text{diag}(\kappa_1(\boldsymbol{q}^\star),\kappa_2(\boldsymbol{q}^\star),\ldots,\kappa_L(\boldsymbol{q}^\star))\boldsymbol{d}$. A direct implication of Lemma 1 concludes that under the sufficiently expensive recourse condition, $\boldsymbol{d}^\star$ defined above is a non-negative linear combination of rows of $\boldsymbol{B}^\star$. Using these notations, we can present the (simplified) dual of problem (17) as

$$\max_{\boldsymbol{\lambda}^\star\in\text{recc}(\boldsymbol{B}^\star)} \sum_{m\in[M]} \theta_m(\boldsymbol{q}^\star)(\boldsymbol{B}^\star_{m:}\boldsymbol{\lambda}^\star) - h(\boldsymbol{q}^\star)(\boldsymbol{d}^{\star\top}\boldsymbol{\lambda}^\star),$$

which is feasible so its objective goes to positive infinity if problem (17) is infeasible. Therefore, we can identify a violated expectation constraint by focusing on the finitely many extreme rays of $\text{recc}(\boldsymbol{B}^\star)$ and for each of them (denoted by $\boldsymbol{\lambda}^\star$), solve the separation problem

$$\max_{\boldsymbol{q}^\star \in \mathcal{Q}_t} \sum_{m\in[M]} \theta_m(\boldsymbol{q}^\star)(\boldsymbol{B}^\star_{m:}\boldsymbol{\lambda}^\star) - h(\boldsymbol{q}^\star)(\boldsymbol{d}^{\star\top}\boldsymbol{\lambda}^\star).$$

6. Numerical Experiments

We conduct three numerical studies to test the performance of the proposed methodology. The first one is a multi-item newsvendor problem (Section 6.1) and the second one is a hospital bed quota allocation problem (Section 6.2), both are two-stage problems. The third study focuses on a fundamental multi-stage single-item inventory control problem (Section 6.3). In our experiments, we evaluate the scalability of our methodology in terms of problem size, and we test two specific examples of infinitely constrained ambiguity sets, namely covariance dominance ambiguity set (in Section 6.1 and Section 6.2) and fourth moment ambiguity set (in Section 6.3). All optimization problems are solved by MOSEK on a 2.3GHz processor with 32GB memory.

6.1. Multi-Item Newsvendor Problem

Consider the newsvendor who faces random demands (Hadley and Whitin 1963). For each item $i \in [I]$, the unit selling price, ordering cost, salvage cost and stock-out cost are denoted as v_i, c_i, g_i and b_i, respectively. We assume $c_i < v_i$ and $g_i < v_i$ to ensure that this problem is profitable without arbitrage opportunity. For any item $i \in [I]$, we denote the order quantity as x_i and the realized demand as z_i; hence the corresponding sale is $\min\{x_i, z_i\}$. The total cost is

$$\begin{aligned} f(\boldsymbol{x},\boldsymbol{z}) &= \sum_{i\in[I]} \left(c_i x_i - v_i \min\{x_i, z_i\} - g_i\left(x_i - \min\{x_i, z_i\}\right) + b_i\left(z_i - \min\{x_i, z_i\}\right)\right) \\ &= (\boldsymbol{c} - \boldsymbol{v} - \boldsymbol{b})^\top \boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{z} + (\boldsymbol{v} + \boldsymbol{b} - \boldsymbol{g})^\top (\boldsymbol{x} - \boldsymbol{z})^+, \end{aligned}$$

Given a known demand distribution $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I)$, the multi-item newsvendor problem is given by

$$\min_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_\mathbb{P}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] = \min_{\boldsymbol{x}\in\mathcal{X}} (\boldsymbol{c} - \boldsymbol{v} - \boldsymbol{b})^\top \boldsymbol{x} + \mathbb{E}_\mathbb{P}[\boldsymbol{b}^\top \tilde{\boldsymbol{z}} + (\boldsymbol{v} + \boldsymbol{b} - \boldsymbol{g})^\top (\boldsymbol{x} - \tilde{\boldsymbol{z}})^+],$$

where $\mathcal{X}$ is a feasible budget set.

In the single-item newsvendor problem without budget constraint, the optimal order quantity is known to be a celebrated critical quantile of the demand distribution. However, possible correlation among multiple items prohibits the extension of this result, because evaluating the expected positive part in the objective function involves multi-dimensional integration and is computationally prohibitive even when the joint demand distribution is known. In addition, estimating the joint demand distribution of items is also statistically challenging. Hence, it is often of interest to investigate the distributionally robust optimization approach to solve the multi-item newsvendor problem (see, *e.g.*, Hanasusanto et al. 2015 and Natarajan and Teo 2017)

$$\min_{\boldsymbol{x}\in\mathcal{X}} (\boldsymbol{c}-\boldsymbol{v}-\boldsymbol{b})^\top \boldsymbol{x} + \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{C}}} \mathbb{E}_{\mathbb{P}}[\boldsymbol{b}^\top \tilde{\boldsymbol{z}} + (\boldsymbol{v}+\boldsymbol{b}-\boldsymbol{g})^\top (\boldsymbol{x}-\tilde{\boldsymbol{z}})^+], \tag{18}$$

which considers complete covariance information of uncertain demands captured in the following covariance dominance ambiguity set in the format of an infinitely constrained ambiguity set:

$$\mathcal{F}_{\mathrm{C}} = \left\{ \mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^I\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}}\sim\mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}[(\boldsymbol{q}^\top(\tilde{\boldsymbol{z}}-\boldsymbol{\mu}))^2] \le \boldsymbol{q}^\top \boldsymbol{\Sigma}\boldsymbol{q} \quad \forall \boldsymbol{q}\in\mathcal{Q} \\ \mathbb{P}[\tilde{\boldsymbol{z}}\in\mathcal{W}] = 1 \end{array} \right\},$$

where $\mathcal{Q} = \{\boldsymbol{q}\in\mathbb{R}^I \mid \|\boldsymbol{q}\|_2 \le 1\}$. An alternative is to replace the collection of infinitely many expectation constraints $\mathbb{E}_{\mathbb{P}}[(\boldsymbol{q}^\top(\tilde{\boldsymbol{z}}-\boldsymbol{\mu}))^2] \le \boldsymbol{q}^\top\boldsymbol{\Sigma}\boldsymbol{q}\ \forall\boldsymbol{q}\in\mathcal{Q}$ by a single conic inequality $\mathbb{E}_{\mathbb{P}}[(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})^\top] \preceq \boldsymbol{\Sigma}$. However, this alternative will typically lead to a reformulation that involves positive semidefinite constraints, which does not scale gracefully and can be very hard to solve when the problem further involves discrete decision variables. In stark contrast, we adopt in our proposed methodology the relaxed ambiguity sets to the above covariance dominance ambiguity set, given by

$$\mathcal{F}_{\mathrm{R}} = \left\{ \mathbb{P}\in\mathcal{P}_0\left(\mathbb{R}^I\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}}\sim\mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}[(\boldsymbol{q}^\top(\tilde{\boldsymbol{z}}-\boldsymbol{\mu}))^2] \le \boldsymbol{q}^\top \boldsymbol{\Sigma}\boldsymbol{q} \quad \forall \boldsymbol{q}\in\bar{\mathcal{Q}} \\ \mathbb{P}[\tilde{\boldsymbol{z}}\in\mathcal{W}] = 1 \end{array} \right\}$$

for some $\bar{\mathcal{Q}}$ with $|\bar{\mathcal{Q}}| < \infty$, which leads to a second-order cone program reformulation that scales gracefully in practice and allows discrete decisions. Commonly used relaxed ambiguity sets include the marignal moment ambiguity set with $\bar{\mathcal{Q}} = \{\boldsymbol{e}_i\}_{i\in[I]}$ (see, *e.g.*, Mak et al. 2014) and the partial cross-moments ambiguity set where $\bar{\mathcal{Q}}$ contains $\{\boldsymbol{e}_i\}_{i\in[I]}$ and some other elements such as $\mathbf{1}$. For adjustable optimization problems, Bertsimas et al. (2019) show for adjustable optimization problems that even the little additional cross-moments information imposed in the later can improve the model performance and raise a question on how to systematically adapt and improve from partial cross-moments (towards complete covariance). As shown in our numerical evidences in this newsvendor problem as well as the other two problems coming subsequently, our proposed GIP algorithm, by identifying violating expectation constraints and improving the ELDR approximation, provides a positive answer to this interesting question.

Introducing an adjustable decision $\boldsymbol{y}$, we can reformulate problem (18) as a two-stage problem:

$$\begin{array}{lll} \min & (\boldsymbol{c}-\boldsymbol{v}-\boldsymbol{b})^\top \boldsymbol{x} + \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{C}}} \mathbb{E}_{\mathbb{P}}[\boldsymbol{b}^\top \tilde{\boldsymbol{z}} + (\boldsymbol{v}+\boldsymbol{b}-\boldsymbol{g})^\top \boldsymbol{y}(\tilde{\boldsymbol{z}})] & \\ \text{s.t.} & \boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{x}-\boldsymbol{z} & \forall \boldsymbol{z}\in\mathcal{W} \\ & \boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{0} & \forall \boldsymbol{z}\in\mathcal{W} \\ & \boldsymbol{x}\in\mathcal{X},\ \boldsymbol{y}\in\mathcal{R}^{I,L}, & \end{array} \tag{19}$$

which can be solved by our framework. Quite notably, the recourse matrix $\boldsymbol{B}$ herein possesses a property of simple recourse that is stronger than complete recourse.

Definition 4 (Simple recourse). The second-stage problem (1) has simple recourse, if and only if it has complete recourse and each row of the recourse matrix $\boldsymbol{B}$ is a standard basis vector.

The simple recourse condition requires that $\boldsymbol{B}_{m:}^\top = \boldsymbol{e}_{v(m)}$ for some $v(m)$-th standard basis vector $\boldsymbol{e}_{v(m)}$, where $v(\cdot)$ is a mapping from the set $[M]$ to the set $[L]$. In such cases, the extreme rays of $\mathrm{recc}(\boldsymbol{B})$ are standard basis vectors and the number of extreme rays is equal to the number of adjustable decisions $\boldsymbol{y}$ (*i.e.*, the number I of items in this experiment). Consequently, the separation problem (15) can be significantly simplified.

Theorem 3. *Suppose the second-stage problem* (1) *has simple recourse, then the indicator problem* (14) *is unbounded for a particular* $\boldsymbol{q}\in\mathcal{Q}\setminus\bar{\mathcal{Q}}$ *if and only if for some* $l\in[L]$*, it holds that* $\sum_{m\in\mathcal{M}_l}\theta_m(\boldsymbol{q}) - h(\boldsymbol{q})d_l > 0$*. Here for each fixed* $l\in[L]$*,* $\mathcal{M}_l = \{m\in[M] : v(m)=l\}$*.*

Proof of Theorem 3. Note that under the simple recourse condition, every extreme ray satisfies $\boldsymbol{\lambda}^\star \geq \boldsymbol{0}$. In addition, for each $m\in[M]$, we have $\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star = \boldsymbol{e}_{v(m)}^\top\boldsymbol{\lambda}^\star = \lambda^\star_{v(m)}$ for some $v(m)\in[L]$. The objective function of the separation problem (15) can be represented as

$$\begin{aligned} \sum_{m\in[M]} \theta_m(\boldsymbol{q})(\boldsymbol{B}_{m:}\boldsymbol{\lambda}^\star) - h(\boldsymbol{q})(\boldsymbol{d}^\top\boldsymbol{\lambda}^\star) &= \sum_{m\in[M]} \theta_m(\boldsymbol{q})\lambda^\star_{v(m)} - \sum_{l\in[L]} h(\boldsymbol{q})d_l\lambda^\star_l \\ &= \sum_{l\in[L]} \lambda^\star_l \left(\sum_{m\in\mathcal{M}_l} \theta_m(\boldsymbol{q}) - h(\boldsymbol{q})d_l \right), \end{aligned}$$

which is additive and positive homogeneous in $\boldsymbol{\lambda}^\star$. Thus, the objective goes to positive infinity if and only if for some $l\in[L]$, it holds that $\sum_{m\in\mathcal{M}_l}\theta_m(\boldsymbol{q}) - h(\boldsymbol{q})d_l > 0$, concluding the proof. □

To test the performance of our proposed GIP algorithm on solving problem (19), we consider a numerical experiment with set-ups that are inspired by Hanasusanto et al. (2015) and Natarajan and Teo (2017). For a fixed number of items, we generate 100 random instances as follows. We first sample the unit selling price $\boldsymbol{v}$ uniformly from $[5,10]^I$, then we set the unit salvage (resp., stock-out) cost to 10% (resp., 25%) of the unit selling price and sample the unit ordering cost uniformly from 50% to 60% of the unit selling price. The mean demand $\boldsymbol{\mu}$ is sampled uniformly from $[5,100]^I$, while the standard deviation $\boldsymbol{\sigma}$ is sampled from independent uniform distributions on $[\boldsymbol{\mu}, 5\boldsymbol{\mu}]$.

The correlations among demands are generated by first sampling a random matrix $\boldsymbol{\Upsilon} \in \mathbb{R}^{I \times I}$ with independent elements uniformly distributed in $[\Delta, 1]$, and then setting the correlation matrix to $\text{diag}(\boldsymbol{w})\boldsymbol{V}\text{diag}(\boldsymbol{w})$, where $\boldsymbol{V} = \boldsymbol{\Upsilon}^\top \boldsymbol{\Upsilon}$ and $\boldsymbol{w}$ is a vector whose i-th element is defined as $w_i = 1/\sqrt{V_{ii}}$. Note that in the above process, we set the parameter Δ to be non-negative, and so the demands are positively correlated; also, a large value of Δ implies that the demands are highly correlated. In particular, we vary the parameter Δ from $\{0, 0.25, 0.5, 0.75\}$. We consider a box-typed support set $\mathcal{W} = \{\boldsymbol{z} \in \mathbb{R}^I_+ \mid \boldsymbol{z} \leq \boldsymbol{\mu} + 3\boldsymbol{\sigma}\}$ with an upper bound scales with both mean and standard deviation. Lastly, we consider the continuous feasible budget set $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^I_+ \mid \mathbf{1}^\top \boldsymbol{x} \leq \boldsymbol{c}^\top (\boldsymbol{\mu} + \boldsymbol{\sigma})\}$.

We start with the marginal moment ambiguity set and improve from that with our GIP algorithm. In this experiment, we set the maximal number of iterations to be 50, and for practical considerations, we also terminate the GIP algorithm when the improvement per iteration is less than 0.2% for five consecutive iterations. We benchmark against the exact solution of problem (18), obtained from solving the equivalent reformulation

$$\min_{\boldsymbol{x} \in \mathcal{X}} (\boldsymbol{c} - \boldsymbol{v} - \boldsymbol{b})^\top \boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{\mu} + \sup_{\mathbb{P} \in \mathcal{F}_{\text{C}}} \mathbb{E}_{\mathbb{P}} \left[\max_{\mathcal{S}: \mathcal{S} \subseteq [I]} \sum_{i \in \mathcal{S}} (v_i + b_i - g_i)(x_i - \tilde{z}_i) \right], \tag{20}$$

which is a special case of problem (5) and can be reformulated as a positive semidefinite program wherein the number of constraints grows linearly with the number 2^I of subsets $\mathcal{S} \subseteq [I]$. Hence, we will first study the cases of 5, 8, and 10 items so that exact optimal solutions are available for comparisons. Average and median relative gaps to the exact objective value among 100 instances are summarized in Table 1. It can be seen that the conservativeness of considering only marginal moment is more obvious as the positive correlations among demands get stronger. On the other hand, by iteratively incorporating covariance information via more partial cross-moments, our GIP algorithm can yield solutions that gradually (and significantly) mitigate the conservativeness. This reveals the importance and benefits of considering covariance information in adjustable distributionally robust optimization problems. The difference between average and median gaps suggests that there are some extreme instances where the ELDR approximation might be inferior. Consequently, the ELDR approximation could perform even better if these outliers were discarded.

We next compare the GIP algorithm and the exact approach in terms of computational scalability. To this end, we consider both continuous and discrete feasible budget sets, given by $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^I_+ \mid \mathbf{1}^\top \boldsymbol{x} \leq \boldsymbol{c}^\top (\boldsymbol{\mu} + \boldsymbol{\sigma})\}$ and $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{Z}^I_+ \mid \mathbf{1}^\top \boldsymbol{x} \leq \boldsymbol{c}^\top (\boldsymbol{\mu} + \boldsymbol{\sigma})\}$, respectively. We fix $\Delta = 0.5$ and for the exact optimal solution to be available within one hour, we study 3 to 13 items when $\boldsymbol{x}$ is continuous as well as 3 to 10 items when $\boldsymbol{x}$ is discrete. It is clear that the GIP algorithm, always achieving average relative gaps that are within 4.5%, pays the cost of extra computation time for iteratively improving the ELDR approximation; see the comparison between columns "MM" and

5 items			8 items			10 items		
Δ	MM	GIP	Δ	MM	GIP	Δ	MM	GIP
0	2.1 [0.5]	0.6 [0.3]	0	44.3 [4.1]	4.4 [1.0]	0	31.6 [8.0]	6.4 [1.5]
0.25	22.6 [2.4]	3.0 [0.6]	0.25	81.3 [33.9]	3.8 [1.4]	0.25	133.2 [74.7]	7.9 [3.3]
0.50	44.9 [16.2]	3.4 [0.6]	0.50	357.7 [74.7]	4.6 [1.2]	0.50	384.3 [149.3]	7.7 [2.8]
0.75	166.2 [48.2]	3.5 [0.1]	0.75	336.3 [128.4]	2.8 [0.4]	0.75	415.5 [176.6]	3.8 [0.9]

Table 1 Average and median (in brackets) relative gaps (%) to the exact objective value among 100 instances: 5 items (left), 8 items (middle) and 10 items (right). Here, "MM" denotes the marginal moment model solved by the initial iteration of GIP algorithm and "GIP" denotes implementing the GIP algorithm for at most 50 iterations.

"GIP" in Table 2. Such extra computation time also makes the GIP algorithm slower than the exact approach when the number of items is small. Nevertheless, the GIP algorithm scales better than the exact approach that solves a positive semidefinite program. As expected, this advantage in computational scalability is more notable when discrete decisions are involved.

6.2. Hospital Quota Allocation Problem

Consider allocating bed quotas for elective admission inpatients to maximize bed utilization (Meng et al. 2015). In this problem, an inpatient can stay in the hospital for at most L days, and we denote the first day of a T-day planning horizon by day 0. The model considers all days in $\mathcal{T} := \mathcal{T}^- \cup \mathcal{T}^+$ where $\mathcal{T}^- = \{1-L, \cdots, -1\}$ and $\mathcal{T}^+ = \{0, \cdots, T-1\}$. In particular, $\mathcal{T}^-$ denotes the days before day 0 on which admitted inpatients are possibly still in the hospital during the planning horizon, and $\mathcal{T}^+$ denotes the days in our planning horizon.

Two types of inpatients are considered in this problem: emergency inpatients (EMIs) and elective admission inpatients (EAIs). It is assumed that EMIs are guaranteed to have beds allocated immediately, and we need to determine the daily bed quota allocated to EAIs. For each $k \in \mathcal{T}^+$ we let x_k be the bed quota allocated to EAIs on day k, and we collectively denote these quotas as $\boldsymbol{x} \in \mathbb{Z}^T$. Daily demands of both types are uncertain (Meng et al. 2015) and they are assumed to be independent of each other. For each $k \in \mathcal{T}^+$ and $l \in [L]$, we let $z_{k,l}$ be the proportion of EAIs who start hospitalization on day k and staying for at least l days; that is, $z_{k,l}x_k$ is the number of EAIs who start hospitalization on day k and staying for at least l days. We use $\xi_{k,l}$ to denote the number of EMIs who stay for at least l days starting from day k.

We consider a cycle of H days (*e.g.*, every week is a cycle if $H = 7$) and we aim to minimize the sum of the maximal bed shortage at every cycle over the planning horizon. This setting is generic and covers several important cases: (i) when $H = T$, we minimize the maximal daily bed shortage

number of items	Computation times			Relative gaps	
	MM	GIP	EXACT	MM	GIP
3	1.6 [1.6]	18.8 [20.4]	1.0 [1.0]	170.1 [40.8]	2.1 [0.0]
5	2.1 [2.0]	71.8 [60.1]	1.3 [1.3]	87.9 [44.9]	1.4 [0.2]
7	2.9 [2.8]	208.2 [159.3]	2.8 [2.8]	280.5 [51.3]	4.4 [1.0]
9	4.0 [3.9]	485.5 [427.9]	13.8 [13.7]	75.5 [57.0]	2.3 [2.0]
11	5.3 [5.1]	1120.7 [856.2]	166.2 [162.9]	99.7 [68.2]	3.5 [3.3]
13	6.9 [6.8]	1945.7 [1266.2]	3454.6 [3446.2]	80.2 [60.4]	4.4 [4.1]

number of items	Computation times			Relative gaps	
	MM	GIP	EXACT	MM	GIP
3	1.9 [1.9]	13.8 [11.7]	0.4 [0.4]	171.5 [41.7]	2.7 [0.0]
5	2.5 [2.5]	70.7 [64.2]	1.5 [1.4]	87.9 [44.9]	1.5 [0.4]
7	3.2 [3.2]	202.2 [154.6]	18.9 [18.3]	273.6 [52.2]	4.3 [1.1]
9	4.4 [4.3]	363.3 [282.8]	391.5 [368.6]	72.5 [55.0]	2.1 [1.4]
10	5.1 [5.1]	667.4 [561.4]	2266.5 [2233.6]	115.3 [68.2]	2.9 [2.2]

Table 2 Average and median (in brackets) of computation times in seconds and relative gaps (%) to the exact objective values among 100 random instances with continuous decisions (upper panel) and with discrete decisions (lower panel). Here, “EXACT” further denotes the exact approach that solves problem (20).

over T days, and (ii) when $H=1$, we minimize the total bed shortages over the planning horizon. For simplicity, we take both T/H and $T/7$ as integers. We consider the objective function

$$f_H(\boldsymbol{x},\boldsymbol{z},\boldsymbol{\xi}) = \sum_{i\in[T/H]} \left\{ \max_{t\in\mathcal{A}_H(i)} \left(\sum_{(k,l)\in\mathcal{U}_t} (z_{k,l}x_k + \xi_{k,l}) - c_t \right) \right\},$$

where for each $t\in\mathcal{T}^+$ and $i\in[T/H]$, c_t is the bed capacity on day t, $\mathcal{A}_H(i) := \{(i-1)H+(j-1) : j\in[H]\}$ contains the days in the i-th cycle, and $\mathcal{U}_t$ is the set of (k,l) pairs such that $(k,l)\in\mathcal{U}_t$ represents that on day t, $z_{k,l}x_k$ beds are needed for the EAIs who are admitted on day k and will stay for at least l days. Note that $\mathcal{U}_t := \{(k,l) : k\in\mathcal{T},\ l\in\mathcal{H}_k,\ l+k=t+1\}$ with $\mathcal{H}_t := \{\max\{1,1-t\},\max\{1,1-t\}+1,\cdots,\min\{L,T-t\}\}$. Assuming that the distribution $\mathbb{P}$ lies in an ambiguity set $\mathcal{F}_{\mathrm{C}}$, we solve the distributionally robust optimization problem

$$Z(H) = \min_{\boldsymbol{x}\in\mathcal{X}} \sup_{\mathbb{P}\in\mathcal{F}_{\mathrm{C}}} \mathbb{E}_{\mathbb{P}}\left[f_H(\boldsymbol{x},\tilde{\boldsymbol{z}},\tilde{\boldsymbol{\xi}}) \right], \tag{21}$$

where $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{Z}^T : \underline{x} \cdot \mathbf{1} \leq \boldsymbol{x} \leq \overline{x} \cdot \mathbf{1}, \sum_{t=0}^{6} x_{(t+7(i-1))} = \hat{x} \; \forall i \in [T/7]\}$ is the feasible set with $\underline{x}$ and $\overline{x}$ being the lower and upper bounds on each daily quota, respectively, while $\hat{x}$ being the weekly quota. Introducing a vector of auxiliary decision variables $\boldsymbol{y}$, problem (21) can be reformulated as the following two-stage problem:

$$Z(H) = \begin{cases} \min & \sup_{\mathbb{P} \in \mathcal{F}_{\mathrm{C}}} \mathbb{E}_{\mathbb{P}}[\mathbf{1}^\top \boldsymbol{y}(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{\xi}})] \\ \text{s.t.} & y_i(\boldsymbol{z}, \boldsymbol{\xi}) \geq \max_{t \in \mathcal{A}_H(i)} \left(\sum_{(k,l) \in \mathcal{U}_t} (z_{k,l} x_k + \xi_{k,l}) - c_t \right) \; \forall (z_{k,l}, \xi_{k,l}) : k \geq 0, \; (\boldsymbol{z}, \boldsymbol{\xi}) \in \mathcal{W}, \; i \in [T/H] \\ & \boldsymbol{x} \in \mathcal{X}, \; \boldsymbol{y} \in \mathcal{R}^{2I, T/H}. \end{cases} \tag{22}$$

We focus on the case of $T = 14$ and $L = 14$, and we study two models $Z(T)$ and $Z(T/2)$ that minimize, respectively, the all-day maximal bed shortage and the sum of weekly maximal bed shortages. The ambiguity set $\mathcal{F}_{\mathrm{C}}$ encompasses the covariance information of $\tilde{\boldsymbol{z}}$ and $\tilde{\boldsymbol{\xi}}$, captured by

$$\mathcal{F}_{\mathrm{C}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{2I}\right) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{\xi}}) \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\nu} \\ \mathbb{E}_{\mathbb{P}}[(\boldsymbol{q}^\top (\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))^2] \leq \boldsymbol{q}^\top \boldsymbol{\Omega} \boldsymbol{q} \quad \forall \boldsymbol{q} \in \mathcal{Q} \\ \mathbb{E}_{\mathbb{P}}[(\boldsymbol{s}^\top (\tilde{\boldsymbol{\xi}} - \boldsymbol{\nu}))^2] \leq \boldsymbol{s}^\top \boldsymbol{\Sigma} \boldsymbol{s} \quad \forall \boldsymbol{s} \in \mathcal{S} \\ \mathbb{P}[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{\xi}}) \in \mathcal{W}] = 1 \end{array} \right\},$$

where $\mathcal{Q} = \mathcal{S} = \{\boldsymbol{q} \in \mathbb{R}^I | \; \|\boldsymbol{q}\|_2 \leq 1\}$. Note that the ambiguity set considered in Meng et al. (2015), consisting of only marginal moment information, is in fact a relaxation of $\mathcal{F}_{\mathrm{C}}$. Define $\mathcal{U} := \cup_{t \in \mathcal{T}^+} \mathcal{U}_t$ as well as two random vectors $\tilde{\boldsymbol{z}} = \{z_{k,l}\}_{(k,l) \in \mathcal{U}}$ and $\tilde{\boldsymbol{\xi}} = \{\xi_{k,l}\}_{(k,l) \in \mathcal{U}}$, it then holds that $I = T(1+T)/2$ if $T \leq L$ and $I = L(1+L)/2 + L(T-L)$ otherwise. For any $(k,l) \in \mathcal{U}$, the upper bounds of $\tilde{z}_{k,l}$ and $\tilde{\xi}_{k,l}$ are $\bar{\mu} = 1$ and $\bar{\nu} = 60$, and the their means are $\mathbb{E}_{\mathbb{P}}[\tilde{z}_{k,l}] = (2L - 2l + 1)\bar{\mu}/(2L)$ and $\mathbb{E}_{\mathbb{P}}[\tilde{\xi}_{k,l}] = (2L - 2l + 1)\bar{\nu}/(2L)$. The support set of $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{\xi}})$ is a polyhedron

$$\mathcal{W} = \left\{ (\boldsymbol{z}, \boldsymbol{\xi}) \in \mathbb{R}^{2I} \;\middle|\; \forall (k,l), \; (k,l') \in \mathcal{U}, \; l > l' : \begin{array}{l} 0 \leq z_{k,l} \leq z_{k,l'} \leq \bar{\mu} \\ 0 \leq \xi_{k,l} \leq \xi_{k,l'} \leq \bar{\nu} \end{array} \right\},$$

and upper bounds on the covariance of $\tilde{\boldsymbol{z}}$ and $\tilde{\boldsymbol{\xi}}$ are $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$, respectively. To specify $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$, we generate the standard deviations and correlations of the random components as follows. For any $(k,l) \in \mathcal{U}$, the standard deviations of $\tilde{z}_{k,l}$ and $\tilde{\xi}_{k,l}$ are $\bar{\mu}/(6L)$ and $\bar{\nu}/(6L)$, respectively. Data of inpatients in different weeks (*e.g.*, day 0 to 6 are in the same week, while day 6 and day 7 are not) are independent, and the random variables of same-type inpatients starting hospitalized in the same week have a correlation matrix generated in the way as in Section 6.1 with $\Delta = 0.25$.

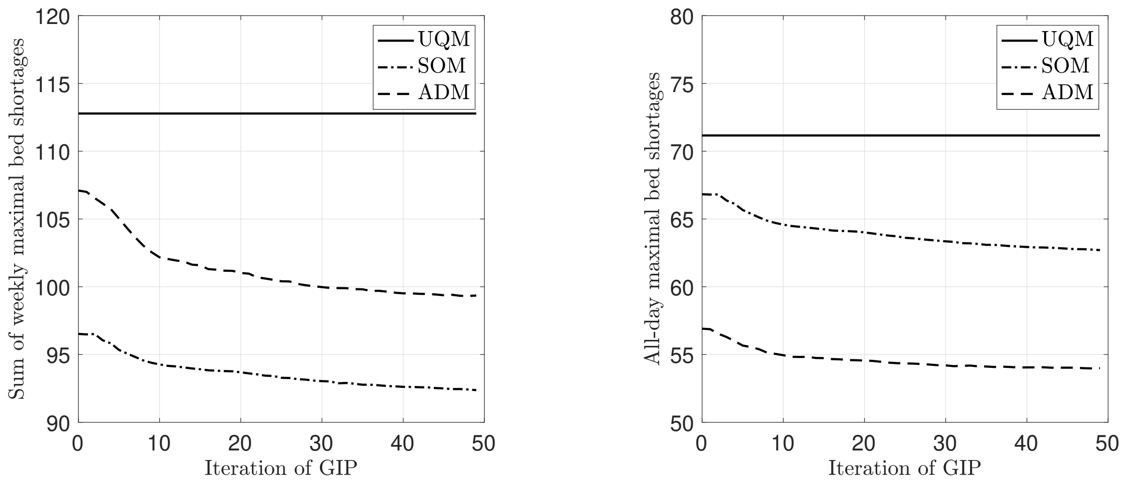

Figure 1 Average bed shortages of different models over 50 random instances and 10000 testing samples: sum of weekly maximal bed shortages (left) and all-day maximal bed shortages (right).

Configurations of the planning horizon are the same as in Meng et al. (2015). For every $t \in \mathcal{T}^+$, the bed capacities are equally $c_t = c = 650$; the weekly quota is $\hat{x} = 301$; and the lower (resp., upper) bound of daily quotas is $\underline{x} = 5$ (resp., $\overline{x} = 80$). The ELDR approximation is applied to solve problem (22): in particular, we start with the marginal moment ambiguity set and use the GIP algorithm to iteratively tighten the relaxed ambiguity set (as well as its lifted counterpart), which further improves the ELDR approximation iteratively. In this experiment, the maximal number of iterations is set to be 50. We randomly generate 50 instances of different $\{(z_{k,l}, \xi_{k,l})\}_{k<0,\, l\in[L]}$[4] with means and covariances as stated above, and for each random instance, we further generate 10000 testing samples of $\{(z_{k,l}, \xi_{k,l})\}_{(k,l)\in\mathcal{U}}$ to evaluate the quality of the quota allocation $\boldsymbol{x}$ over the whole planning horizon, for three models as follows.

- **Uniform quota model (UQM)**: the weekly quota are equally allocated to days in a week.
- **Sum of weekly maximums (SOM)**: we solve an instance of problem (22), $Z(7)$.
- **All-Day Maximum (ADM)**: we solve an instance of problem (22), $Z(14)$.[5]

For all random instances, the computation times of the SOM and ADM models are similar and are both about three hours, certifying the two models as efficient tools for scheduling weekly bed allocation. As for the performance, it is illustrated in Figure 1 and Table 3 that, by applying the corresponding quota allocation strategies, SOM and ADM achieve the smallest sum of weekly maximal bed shortages (criterion 1) and all-day maximal bed shortage (criterion 2), respectively. This

[4] In problem (22), $\{(z_{k,l}, \xi_{k,l})\}_{k<0,\, l\in[L]}$ are constant because we already know the inpatients admitted on days before the planning horizon (*i.e.*, the inpatients admitted in $\mathcal{T}^-$).

[5] The optimized robust model in Meng et al. (2015) is essentially an ADM model with only marginal moment.

Criterion 1			
Number of Iterations	UQM	SOM	ADM
0	112.8	96.5	107.1
10	112.8 [0%]	94.3 [2.3%]	102.2 [4.6%]
20	112.8 [0%]	93.7 [2.9%]	101.0 [5.7%]
30	112.8 [0%]	93.0 [3.6%]	100.0 [6.7%]
40	112.8 [0%]	92.6 [4.1%]	99.5 [7.1%]
50	112.8 [0%]	92.4 [4.3%]	99.3 [7.3%]

Criterion 2			
Number of Iterations	UQM	SOM	ADM
0	71.2	66.8	56.9
10	71.2 [0%]	64.6 [3.4%]	55.0 [3.4%]
20	71.2 [0%]	64.0 [4.2%]	54.6 [4.1%]
30	71.2 [0%]	63.3 [5.2%]	54.2 [4.8%]
40	71.2 [0%]	62.9 [5.8%]	54.0 [5.0%]
50	71.2 [0%]	62.7 [6.2%]	54.0 [5.1%]

Table 3 Sum of weekly maximal bed shortages (left) and maximal bed shortage (right) (percentage of bed shortage decreases compared with iteration 0 in brackets) of all days of the three models.

implies that one should choose a different value of H for the model $Z(H)$ with a different optimization objective. Observe that UQM always gives the worst performance among the three models: bed shortages under criterion 1 (resp., criterion 2) are over 110 (resp., 70) at all time—a notably worse performance compared to the other two adjustable distributionally robust optimization models.

We next turn to the "best" models with respect to the two criteria: after 50 iterations of GIP, SOM improves its performance by 4.3% while ADM improves by 5.1%. This indicates that starting from the marginal moment ambiguity set (*i.e.*, the ambiguity set considered in Meng et al. 2015), the out-of-sample performance of the adjustable distributionally robust optimization models is monotonically improved by iteratively incorporating covariance information. Note that both models tend to improve mildly in the later iterations as the GIP algorithm may converge after some iterations. Indeed, improvement of the last 10 iterations of SOM under criterion 1 (resp., ADM under criterion 2) is merely 0.2% (resp., 0.1%); see the left (resp., right) panel of Table 3.

6.3. Multi-Stage Inventory Control Problem

Consider a T-stage inventory control problem where the uncertain demand in stage t is $\tilde{d}_t$. At the beginning of each stage $t \in [T]$, the order quantity $x_t \in [0, \bar{x}_t]$ is assumed to arrive immediately to replenish the stock before demand realization, and the unit ordering cost, holding cost of excessive inventory and backlogged cost are c_t, h_t and b_t, respectively. The demand process is motivated by Graves (1999) and See and Sim (2010): $\tilde{d}_t = d_t(\tilde{\boldsymbol{z}}_t) = \tilde{z}_t + \alpha\tilde{z}_{t-1} + \cdots + \alpha\tilde{z}_1 + \mu$, where the uncertain factors z_t, $t \in [T]$ are realized periodically and are identically distributed in $[-\bar{z}, \bar{z}]$ with zero mean. For any $t \in [T]$, let $\boldsymbol{z}_t = (z_1, \ldots, z_t)$ for ease of exposition. In particular, we focus on a fourth moment ambiguity set that captures both the covariance and fourth moment information of $\tilde{\boldsymbol{z}}$:

$$\mathcal{F}_{\mathrm{F}} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^T\right) \;\middle|\; \begin{array}{ll} \tilde{\boldsymbol{z}} \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{0} & \\ \mathbb{E}_{\mathbb{P}}[(\boldsymbol{q}^\top \tilde{\boldsymbol{z}})^2] \leq \boldsymbol{q}^\top \boldsymbol{\Sigma} \boldsymbol{q} & \forall \boldsymbol{q} \in \mathcal{Q} \\ \mathbb{E}_{\mathbb{P}}[(\boldsymbol{s}^\top \tilde{\boldsymbol{z}})^4] \leq \varphi(\boldsymbol{s}) & \forall \boldsymbol{s} \in \mathcal{S} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in [-\bar{z}, \bar{z}]^T\right] = 1 & \end{array} \right\},$$

where $\mathcal{Q} = \mathcal{S} = \{\boldsymbol{q} \in \mathbb{R}^T |\ \|\boldsymbol{q}\|_2 \leq 1\}$. The diagonal matrix $\boldsymbol{\Sigma}$ is given by $\Sigma_{tt} = \bar{z}^2/3$ for all $t \in [T]$ and $\varphi(\boldsymbol{s}) = \sum_{t \in [T]} (s_t^4 \tau_t^4 + 6 \sum_{r > t,\, r \in [T]} s_t^2 s_r^2 \Sigma_{tt} \Sigma_{rr})$ imposes an upper bound on the fourth moment of $\boldsymbol{s}^\top \tilde{\boldsymbol{z}}$. Here, for all $t \in [T]$, $\tau_t^4 = \bar{z}^4/18$ is the upper bound on the fourth moment of $\tilde{z}_t$. The objective is to minimize the worst-case expected total cost over the entire planning horizon:

$$\begin{array}{lll} \min & \sup\limits_{\mathbb{P} \in \mathcal{F}_{\mathrm{F}}} \mathbb{E}_{\mathbb{P}} \left[\sum\limits_{t=1}^{T} \left(c_t x_t(\tilde{\boldsymbol{z}}_{t-1}) + y_t(\tilde{\boldsymbol{z}}_t) \right) \right] & \\ \text{s.t.} & y_t(\boldsymbol{z}_t) \geq b_t \left(\sum\limits_{v=1}^{t} (d_v(\boldsymbol{z}_v) - x_v(\boldsymbol{z}_{v-1})) \right) & \forall \boldsymbol{z} \in [-\bar{z}, \bar{z}]^T,\ t \in [T] \\ & y_t(\boldsymbol{z}_t) \geq h_t \left(\sum\limits_{v=1}^{t} (x_v(\boldsymbol{z}_{v-1}) - d_v(\boldsymbol{z}_v)) \right) & \forall \boldsymbol{z} \in [-\bar{z}, \bar{z}]^T,\ t \in [T] \\ & 0 \leq x_t(\boldsymbol{z}_{t-1}) \leq \bar{x}_t & \forall \boldsymbol{z} \in [-\bar{z}, \bar{z}]^T,\ t \in [T] \\ & x_t \in \mathcal{R}^{t-1,1},\ y_t \in \mathcal{R}^{t,1} & \forall t \in [T]. \end{array} \tag{23}$$

Quite notably, the number of extreme rays of the recession cone generated by the recourse matrix in problem (23) is identical to the number of stages.

THEOREM 4. *Consider a finite horizon, T-stage inventory control problem* (23)*, the recession cone generated by the recourse matrix has T extreme rays.*

Proof of Theorem 4. The first-stage decision in problem (23) is x_1 while the adjustable decisions are x_t, $t = 2, \ldots, T$ and y_t, $t = 1, \ldots, T$. We can then represent the constraints as $\hat{\boldsymbol{a}}(\boldsymbol{z}) x_1 + \boldsymbol{B}(x_2, \ldots, x_T, y_1, \ldots, y_T) \geq \boldsymbol{b}(\boldsymbol{z})$, where $\hat{\boldsymbol{a}}(\boldsymbol{z}) = (0, 0, \ldots, 0, 0, b_1, -h_1, \ldots, b_T, -h_T) \in \mathbb{R}^{4T-2}$,

$$\boldsymbol{b}(\boldsymbol{z}) = \left(0, -\bar{x}_2, \ldots, 0, -\bar{x}_T, b_1 d_1(z_1), -h_1 d_1(z_1), \ldots, b_T \sum_{v=1}^{T} d_v(\boldsymbol{z}_v), -h_T \sum_{v=1}^{T} d_v(\boldsymbol{z}_v) \right) \in \mathbb{R}^{4T-2}$$

and the recourse matrix $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{O} \\ \boldsymbol{B}_2 & \boldsymbol{B}_3 \end{bmatrix} \in \mathbb{R}^{(4T-2)\times(2T-1)}$ consisting of a zero matrix $\boldsymbol{O} \in \mathbb{R}^{2(T-1)\times T}$, $\boldsymbol{B}_1 \in \mathbb{R}^{2(T-1)\times(T-1)}, \boldsymbol{B}_2 \in \mathbb{R}^{2T\times(T-1)}$ and $\boldsymbol{B}_3 \in \mathbb{R}^{2T\times T}$ such that:

$$\boldsymbol{B}_1 = \begin{bmatrix} 1 & 0 & \cdots\cdots & & 0 \\ -1 & 0 & \cdots\cdots & & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots\cdots & & 0 & 1 \\ 0 & \cdots\cdots & & 0 & -1 \end{bmatrix}, \ \boldsymbol{B}_2 = \begin{bmatrix} 0 & 0 & \cdots\cdots & & 0 \\ 0 & 0 & \cdots\cdots & & 0 \\ b_2 & 0 & 0 & \cdots & 0 \\ -h_2 & 0 & 0 & \cdots & 0 \\ b_3 & b_3 & 0 & \cdots & 0 \\ -h_3 & -h_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_T & b_T & \cdots\cdots & & b_T \\ -h_T & -h_T & \cdots\cdots & & -h_T \end{bmatrix} \ \text{and} \ \boldsymbol{B}_3 = \begin{bmatrix} 1 & 0 & \cdots\cdots & & 0 \\ 1 & 0 & \cdots\cdots & & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots\cdots & & 0 & 1 \\ 0 & \cdots\cdots & & 0 & 1 \end{bmatrix}.$$

By the definition of an extreme ray, in the linear system $\boldsymbol{B\lambda} \geq \boldsymbol{0}$, there are $2T-2$ linearly independent constraints active at the extreme ray $\boldsymbol{\lambda}^\star$. It is clear that for all $i \in [T-1]$, $\lambda_i^\star = 0$. As a consequence, there must be $(T-1)$ constraints active at the last $2T$ constraints. That is to say, the extreme ray $\boldsymbol{\lambda}^\star$ has $(T-1)$ zeros among its last T components. In particular, it must take the form $\boldsymbol{\lambda}^\star = (\boldsymbol{0}, \boldsymbol{e}_i)$ for some i-th standard unit basis in $\mathbb{R}^T$. Hence, there are T extreme rays. □

Theorem 4 reveals that the recession cone generated by the recourse matrix has a number of extreme rays that scales linearly with the number of stages, so does the recession cone generated by any sub-matrix of the recourse matrix that corresponds to a particular $\boldsymbol{q}^\star$ (or $\boldsymbol{s}^\star$)—a direct implication. This is an attractive feature for implementing the GIP algorithm. To solve problem (23), we consider the ELDR approximation with relaxed ambiguity sets as follows, each of which replaces $\mathcal{Q}$ and $\mathcal{S}$ in $\mathcal{F}_{\mathrm{C}}$ with some sets $\bar{\mathcal{Q}}$ and $\bar{\mathcal{S}}$ of finite elements.

(i) **Marginal moment (MM)**: $\bar{\mathcal{S}} = \bar{\mathcal{Q}} = \{\boldsymbol{e}_t\}_{t\in[T]}$.

(ii) **Partial cross-moments (PCM)**: $\bar{\mathcal{S}} = \{\boldsymbol{e}_t\}_{t\in[T]}$ and $\bar{\mathcal{Q}} = \{\boldsymbol{q}_{st} \mid s \leq t, \ t \in [T]\}$ with

$$\boldsymbol{q}_{st} = (\underbrace{0,\ldots,0}_{s-1}, \underbrace{1,\ldots,1}_{t-s+1}, \underbrace{0,\ldots,0}_{T-t})$$

that specifies an upper bound on the variance of $\sum_{r=s}^{t} \tilde{z}_r$; see Bertsimas et al. (2019).

(iii) **GIP**: starting with the ambiguity set in (ii) and implementing the GIP algorithm for identifying $\boldsymbol{q} \in \mathcal{Q} \backslash \bar{\mathcal{Q}}$.

(iv) **GIP$^+$**: starting with the ambiguity set obtained from (iii) and implementing the GIP algorithm for identifying $\boldsymbol{s} \in \mathcal{S} \backslash \bar{\mathcal{S}}$.

For the GIP algorithm described in (*iii*) and (*iv*), the maximal number of iterations is 25, and we may also terminate when the improvement per iteration is less than 5% for 10 consecutive iterations. We study the cases of $T \in \{5, 10, 20\}$ and we set $\bar{x}_t = 260, c_t = 0.1, h_t = 0.02$ for all $t \in [T]$, $b_t = h_t b/h$ for all $t \in [T-1]$ and $b_T = 10b_{T-1}$. Because unfulfilled demands at the last stage are lost, we set the unit backlogged cost relatively high. For the case of $T = 5$, we set $\mu = 200$ and $\bar{z} = 40$, while for $T = 10$ and $T = 20$, we set $\mu = 200$, $\bar{z} = 20$ and $\mu = 240$, $\bar{z} = 12$, respectively.

In Table 4, we present the objective values of different approaches. Observe that when problem size (*i.e.*, T) becomes larger and when α increases, the differences in objective values among different approaches become more significant. The benefit of incorporating more precise covariance information and that of additional fourth moment information are both quite notable. Computation times in Table 5 show the extra cost of these benefits in both the GIP and GIP$^+$ models, which also tends to be higher for larger values of T and α. This is because that cases of large T and α tend to incur large numbers of iterations as well as more computation time per iteration. Note that the GIP$^+$ model is slower than the GIP model because the former needs more time per iteration as it inherits $\bar{\mathcal{Q}}$ from the last iteration of the latter while iteratively enlarging the set $\bar{\mathcal{S}}$. In conclusion, results in Tables 4 and 5 caution us the tradeoff of our approach between improving the ELDR approximation and the extra cost in computation time, which shall be well balanced in practice.

7. Conclusion

We study adjustable distributionally robust optimization problems with an infinitely constrained ambiguity set, and extended upon the linear decision rule technique, we propose an algorithm to obtain approximate solutions that could be monotonically improved in each iteration. We apply our framework to three different applications with the covariance dominance ambiguity set as well as a fourth moment ambiguity set. The numerical results show encouraging benefits of our approach.

Yet, our framework can be directly applied to adjustable distributionally robust optimization problems with a special infinitely constrained ambiguity set called *entropic dominance ambiguity set* that is proposed by Chen et al. (2019), which (*i*) characterizes some interesting properties of the uncertainty such as independence among uncertain components and (*ii*) typically leads to an exponential cone reformulation that can be efficiently solved by the state-of-the-art commercial solvers such as Gurobi and MOSEK. Recently, Chen et al. (2021a) propose a *static* distributionally robust optimization model with the entropic dominance ambiguity set for scheduling electric vehicle charging. The authors point out that in practice, this charging schedule can be adjustable as uncertainty is revealed over time, thus an *adjustable* two-stage model has the potential to inform better decisions. We leave this promising avenue for future research.

	$T=5$				$T=10$				$T=20$			
b/h	MM	PCM	GIP	GIP$^+$	MM	PCM	GIP	GIP$^+$	MM	PCM	GIP	GIP$^+$
	$\alpha=0$				$\alpha=0$				$\alpha=0$			
10	107.5	107.5	–	–	205.5	205.5	–	–	485.0	484.8	–	–
30	108.0	108.0	–	–	206.0	206.0	–	–	486.0	486.0	–	–
50	108.0	108.0	–	–	206.0	206.0	–	–	486.0	486.0	–	–
	$\alpha=0.25$				$\alpha=0.25$				$\alpha=0.25$			
10	108.5	108.5	–	–	205.6	205.6	–	–	694.8	538.4	537.1	506.8
30	109.2	109.2	–	–	206.1	206.1	–	–	917.5	599.0	596.6	510.8
50	109.2	109.2	–	–	206.1	206.1	–	–	1121.6	647.1	645.5	510.1
	$\alpha=0.50$				$\alpha=0.50$				$\alpha=0.50$			
10	134.8	123.7	123.5	117.9	225.1	217.1	216.9	213.4	1072.0	640.6	631.7	564.7
30	164.8	151.8	151.7	127.4	239.2	224.9	224.8	217.1	1974.1	837.8	811.2	592.2
50	191.4	178.6	178.5	136.9	251.9	231.9	231.8	218.9	2876.1	1017.0	965.0	603.0
	$\alpha=0.75$				$\alpha=0.75$				$\alpha=0.75$			
10	168.9	144.7	143.6	133.7	303.4	245.2	244.4	234.1	1497.0	759.9	737.0	661.4
30	233.8	207.7	204.0	160.2	399.3	296.1	293.5	249.9	3214.3	1140.3	1071.8	710.9
50	290.9	266.8	261.5	180.7	481.7	343.0	338.7	260.3	4931.7	1502.4	1016.3	838.7
	$\alpha=1$				$\alpha=1$				$\alpha=1$			
10	205.5	169.9	166.8	152.9	397.3	283.8	280.8	263.3	1939.2	891.6	854.5	779.4
30	320.0	270.8	263.7	189.7	592.8	390.6	380.9	299.3	4505.1	1491.7	1361.2	985.7
50	434.1	361.9	353.3	221.6	769.5	491.5	474.5	327.5	7071.1	2079.0	1854.9	1187.6

Table 4 Performance of the ELDR approximation under different relaxed ambiguity sets: $T=5$ (left), $T=10$ (middle) and $T=20$ (right). Here, the symbol '–' denotes that the GIP model does not improve over the PCM model or the GIP$^+$ model does not improve over the GIP model.

b/h	$T=5$	$T=10$	$T=20$
	$\alpha=0$		
10	10.4	54.1	892.5
30	8.7	56.2	955.4
50	8.7	55.8	941.3
	$\alpha=0.25$		
10	10.1	54.2	7598.6
30	8.6	56.4	8589.3
50	8.6	56.3	6922.4
	$\alpha=0.50$		
10	29.5	232.9	15253.5
30	17.9	177.1	20108.6
50	17.7	114.3	46672.6
	$\alpha=0.75$		
10	64.3	392.0	26410.2
30	73.7	396.2	26451.6
50	68.8	411.1	71067.8
	$\alpha=1$		
10	98.4	431.3	31845.5
30	77.6	692.8	79430.8
50	72.4	710.3	86891.9

b/h	$T=5$	$T=10$	$T=20$
	$\alpha=0$		
10	22.5	117.8	1191.1
30	24.2	114.8	1277.7
50	23.9	114.6	1467.4
	$\alpha=0.25$		
10	263.9	926.2	14282.5
30	25.9	924.3	22521.6
50	24.6	899.2	28585.6
	$\alpha=0.50$		
10	352.1	929.3	26143.8
30	320.5	1060.2	52578.6
50	57.2	1160.1	84172.7
	$\alpha=0.75$		
10	393.8	994.6	50821.3
30	53.3	1288.9	235465.5
50	56.1	1226.4	137633.9
	$\alpha=1$		
10	652.2	1361.8	52578.6
30	346.2	1410.0	212366.4
50	501.2	1422.5	257955.3

Table 5 Computation times in seconds of the GIP model (left) and the GIP^+ model (right).

Acknowledgments

The authors wish to express their gratitude to the editors and two anonymous reviewers for constructive and valuable comments that led to substantial improvements of the paper. This work is supported by the Early Career Scheme from the Hong Kong Research Grants Council (Project No. CityU 21502820), the CityU Start-Up Grant (Project No. 9610481), the CityU Strategic Research Grant (Project No. 7005688), the National Natural Science Foundation of China (Project No. 72032005), and Chow Sang Sang Group Research Fund sponsored by Chow Sang Sang Holdings International Limited (Project No. 9229076).

References

Anderson, Brian, John Moore. 2007. *Optimal control: linear quadratic methods*. Courier Corporation.

Atamtürk, Alper, Muhong Zhang. 2007. Two-stage robust network flow and design under demand uncertainty. *Operations Research* **55**(4) 662–673.

Bandi, Chaithanya, Eojin Han, Omid Nohadani. 2019. Sustainable inventory with robust periodic-affine policies and application to medical supply chains. *Management Science* **65**(10) 4636–4655.

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

Ben-Tal, Aharon, Boaz Golany, Arcadi Nemirovski, Jean-Philippe Vial. 2005. Supplier-retailer flexible commitments contracts: a robust optimization approach. *Manufacturing & Service Operations Management* **7**(3) 589–605.

Ben-Tal, Aharon, Alexander Goryashko, Elana Guslitzer, Arkadi Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming* **99**(2) 351–376.

Ben-Tal, Aharon, Arkadi Nemirovski. 1998. Robust convex optimization. *Mathematics of Operations Research* **23**(4) 769–805.

Bertsimas, Dimitris, Vineet Goyal. 2012. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical Programming* **134**(2) 491–531.

Bertsimas, Dimitris, Vishal Gupta, Nathan Kallus. 2018a. Data-driven robust optimization. *Mathematical Programming* **167**(2) 235–292.

Bertsimas, Dimitris, Vishal Gupta, Nathan Kallus. 2018b. Robust sample average approximation. *Mathematical Programming* **171**(1) 217–282.

Bertsimas, Dimitris, Dan Iancu, Pablo Parrilo. 2010. Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research* **35**(2) 363–394.

Bertsimas, Dimitris, Shimrit Shtern, Bradley Sturt. 2021. Two-stage sample robust optimization. Forthcoming at *Operations Research*.

Bertsimas, Dimitris, Melvyn Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.

Bertsimas, Dimitris, Melvyn Sim, Meilin Zhang. 2019. Adaptive distributionally robust optimization. *Management Science* **65**(2) 604–618.

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

Chen, Li, Long He, Yangfang Helen Zhou. 2021a. An exponential cone programming approach for managing electric vehicle charging. Available at *SSRN 3548028*.

Chen, Xi, Simai He, Bo Jiang, Christopher Ryan, Teng Zhang. 2021b. The discrete moment problem with nonconvex shape constraints. *Operations Research* **69**(1) 279–296.

Chen, Xin, Melvyn Sim, Peng Sun. 2007. A robust optimization perspective on stochastic programming. *Operations Research* **55**(6) 1058–1071.

Chen, Xin, Yuhan Zhang. 2009. Uncertain linear programs: extended affinely adjustable robust counterparts. *Operations Research* **57**(6) 1469–1482.

Chen, Zhi, Melvyn Sim, Peng Xiong. 2020. Robust stochastic optimization made easy with RSOME. *Management Science* **66**(8) 3329–3339.

Chen, Zhi, Melvyn Sim, Huan Xu. 2019. Distributionally robust optimization with infinitely constrained ambiguity sets. *Operations Research* **67**(5) 1328–1344.

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

Duchi, John, Peter Glynn, Hongseok Namkoong. 2021. Statistics of robust optimization: a generalized empirical likelihood approach. Forthcoming at *Mathematics of Operations Research*.

Gao, Rui, Anton Kleywegt. 2016. Distributionally robust stochastic optimization with Wasserstein distance. *arXiv preprint arXiv:1604.02199*.

Garstka, Stanley, Roger Wets. 1974. On decision rules in stochastic programming. *Mathematical Programming* **7**(1) 117–143.

Georghiou, Angelos, Angelos Tsoukalas, Wolfram Wiesemann. 2021. On the optimality of affine decision rules in robust and distributionally robust optimization. Available at *Optimization Online*.

Georghiou, Angelos, Wolfram Wiesemann, Daniel Kuhn. 2015. Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming* **152**(1-2) 301–338.

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

Gounaris, Chrysanthos, Wolfram Wiesemann, Christodoulos Floudas. 2013. The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research* **61**(3) 677–693.

Graves, Stephen. 1999. A single-item inventory model for a nonstationary demand process. *Manufacturing & Service Operations Management* **1**(1) 50–61.

Gupta, Vishal. 2019. Near-optimal Bayesian ambiguity sets for distributionally robust optimization. *Management Science* **65**(9) 4242–4260.

Hadley, George, Thomson Whitin. 1963. *Analysis of inventory systems*. Prentice Hall.

Hanasusanto, Grani, Daniel Kuhn, Stein Wallace, Steve Zymler. 2015. Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming* **152**(1-2) 1–32.

He, Long, Zhenyu Hu, Meilin Zhang. 2020. Robust repositioning for vehicle sharing. *Manufacturing & Service Operations Management* **22**(2) 241–256.

Iancu, Dan, Mayank Sharma, Maxim Sviridenko. 2013. Supermodularity and affine policies in dynamic robust optimization. *Operations Research* **61**(4) 941–956.

Li, Bowen, Ruiwei Jiang, Johanna Mathieu. 2019. Ambiguous risk constraints with moment and unimodality information. *Mathematical Programming* **173**(1) 151–192.

Lorca, Álvaro, Andy Sun, Eugene Litvinov, Tongxin Zheng. 2016. Multistage adaptive robust optimization for the unit commitment problem. *Operations Research* **64**(1) 32–51.

Lu, Mengshi, Zuo-Jun Max Shen. 2021. A review of robust operations management under model uncertainty. *Production and Operations Management* **30**(6) 1927–1943.

Mak, Ho-Yin, Ying Rong, Jiawei Zhang. 2014. Appointment scheduling with limited distributional information. *Management Science* **61**(2) 316–334.

Meng, Fanwen, Jin Qi, Meilin Zhang, James Ang, Singfat Chu, Melvyn Sim. 2015. A robust optimization model for managing elective admission in a public hospital. *Operations Research* **63**(6) 1452–1467.

Mohajerin Esfahani, Peyman, Daniel Kuhn. 2018. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming* **171**(1) 115–166.

Natarajan, Karthik, Chung-Piaw Teo. 2017. On reduced semidefinite programs for second order moment bounds with applications. *Mathematical Programming* **161**(1-2) 487–518.

See, Chuen-Teck, Melvyn Sim. 2010. Robust approximation to multiperiod inventory management. *Operations Research* **58**(3) 583–594.

Shapiro, Alexander, Arkadi Nemirovski. 2005. On complexity of stochastic programming problems. *Continuous Optimization*. Springer, 111–146.

Wang, Zizhuo, Peter Glynn, Yinyu Ye. 2016. Likelihood robust optimization for data-driven problems. *Computational Management Science* **13**(2) 241–261.

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

Zhen, Jianzhe, Dick den Hertog, Melvyn Sim. 2018. Adjustable robust optimization via Fourier–Motzkin elimination. *Operations Research* **66**(4) 1086–1100.