Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Adjustable Robust Optimization via Fourier-Motzkin Elimination

Jianzhe Zhen, Dick den Hertog, Melvyn Sim

informs
http://pubsonline.informs.org/journal/opre/

OPERATIONS RESEARCH
Vol. 66, No. 4, July–August 2018, pp. 1086–1100
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Adjustable Robust Optimization via Fourier–Motzkin Elimination

Jianzhe Zhen,[a] Dick den Hertog,[a] Melvyn Sim[b]

[a] Department of Econometrics and Operations Research, Tilburg University, 5000 LE Tilburg, Netherlands; [b] NUS Business School, National University of Singapore, Singapore 119245

Contact: j.zhen@tilburguniversity.edu, http://orcid.org/0000-0001-9826-2209 (JZ); d.denhertog@tilburguniversity.edu, http://orcid.org/0000-0002-1829-855X (DdH); melvynsim@nus.edu.sg, http://orcid.org/0000-0001-9798-2482 (MS)

Received: November 1, 2016
Revised: June 22, 2017; October 6, 2017
Accepted: November 15, 2017
Published Online in Articles in Advance: July 24, 2018

Subject Classifications: programming: nonlinear: algorithms, theory
Area of Review: Optimization

https://doi.org/10.1287/opre.2017.1714

Abstract. We demonstrate how adjustable robust optimization (ARO) problems with fixed recourse can be cast as static robust optimization problems via Fourier–Motzkin elimination (FME). Through the lens of FME, we characterize the structures of the optimal decision rules for a broad class of ARO problems. A scheme based on a blending of classical FME and a simple linear programming technique that can efficiently remove redundant constraints is developed to reformulate ARO problems. This generic reformulation technique enhances the classical approximation scheme via decision rules, and it enables us to solve adjustable optimization problems to optimality. We show via numerical experiments that, for small-sized ARO problems, our novel approach finds the optimal solution. For moderate- or large-sized instances, we eliminate a subset of the adjustable variables, which improves the solutions obtained from linear decision rules.

Funding: The research of the first author is supported by the Netherlands Organisation for Scientific Research [NWO Grant 613.001.208]. The third author acknowledges the funding support from the Singapore Ministry of Education [Social Science Research Thematic Grant MOE2016-SSRTG-059].

Keywords: Fourier–Motzkin elimination • adjustable robust optimization • linear decision rules • redundant constraint identification

1. Introduction

In recent years, robust optimization has been experiencing an explosive growth and has now become one of the dominant approaches to address decision making under uncertainty. In robust optimization, uncertainty is described by a distribution-free uncertainty set, which is typically a conic representable bounded convex set (see, for instance, El Ghaoui and Lebret 1997; El Ghaoui et al. 1998; Ben-Tal and Nemirovski 1998, 1999, 2000; Bertsimas and Sim 2004; Bertsimas and Brown 2009; Bertsimas et al. 2011). Among other benefits, robust optimization offers a computationally viable methodology for immunizing mathematical optimization models against parameter uncertainty by replacing probability distributions with uncertainty sets as fundamental primitives. It has been successful in providing computationally scalable methods for a wide variety of optimization problems.

The seminal work Ben-Tal et al. (2004) extends classical robust optimization to encompass adjustable decisions. Adjustable robust optimization (ARO) is a methodology to help decision makers make robust and resilient decisions that extend well into the future. In contrast to robust optimization, some of the decisions in ARO problems can be adjusted at a later moment in time after (part of) the uncertain parameter has been revealed. ARO yields less conservative decisions than robust optimization, but ARO problems are in general computationally intractable. To circumvent the intractability, Ben-Tal et al. (2004) restrict the adjustable decisions to be affinely dependent on the uncertain parameters, an approach known as linear decision rules (LDRs).

Bertsimas et al. (2010), Iancu et al. (2013) and Bertsimas and Goyal (2012) establish the optimality of LDRs for some important classes of ARO problems. Chen and Zhang (2009) further improve LDRs by extending the affine dependency to the auxiliary variables that are used in describing the uncertainty set. Henceforth, variants of piecewise affine decision rules have been proposed to improve the approximation while maintaining the tractability of the adjustable distributionally robust optimization (ADRO) models. Such approaches include the deflected and segregated LDRs of Chen et al. (2008), the truncated LDRs of See and Sim (2009), and the bideflected and (generalized) segregated LDRs of Goh and Sim (2010). In fact, LDRs were discussed in the early literature of stochastic programming, but the technique had been abandoned as a result of suboptimality (see Garstka and Wets 1974). Interestingly, there is also a revival of using LDRs for solving multistage stochastic optimization problems (Kuhn et al. 2011). Other nonlinear decision rules in the recent literature include, for example, quadratic decision rules in Ben-Tal et al. (2009) and polynomial decision rules in Bertsimas et al. (2011).

Another approach for ARO problems is finite adaptability, in which the uncertainty set is split into a number of smaller subsets, each with its own set of recourse decisions. The number of these subsets can be either fixed a priori or decided by the optimization model (Vayanos et al. 2011, Bertsimas and Caramanis 2010, Hanasusanto et al. 2014, Postek and den Hertog 2016, Bertsimas and Dunning 2016).

It has been observed that robust optimization models can lead to an underspecification of uncertainty because they do not exploit distributional knowledge that may be available. In such cases, (adjustable) robust optimization may propose overly conservative decisions. In the era of modern business analytics, one of the biggest challenges in operations research concerns the development of highly scalable optimization problems that can accommodate vast amounts of noisy and incomplete data while at the same time truthfully capturing the decision maker's attitude toward risk (exposure to uncertain outcomes whose probability distribution is known) and ambiguity (exposure to uncertainty about the probability distribution of the outcomes). One way of dealing with risk is via stochastic programming. These methods assume that the underlying distribution of the uncertain parameter is known, but they do not incorporate ambiguity in their decision criteria for optimization. For references on these techniques, we refer to Birge and Louveaux (1997) and Kali and Wallace (1994). In evaluating preferences over risk and ambiguity, Scarf (1958) is the first to study a single-product newsvendor problem where the precise demand distribution is unknown but is only characterized by its mean and variance. Subsequently, such models have been extended to minimax stochastic optimization models (see, for instance, Žáčková 1966, Breton and El Hachem 1995, Shapiro and Kleywegt 2002, Shapiro and Ahmed 2004) and recently to distributionally robust optimization models (see, for instance, Chen et al. 2007, Chen and Sim 2009, Popescu 2007, Delage and Ye 2010, Xu and Mannor 2012). In terms of tractable formulations for a wide variety of static robust convex optimization problems, Wiesemann et al. (2014) propose a broad class of ambiguity sets where the family of probability distributions are characterized by conic representable expectation constraints and nested conic representable confidence sets. Chen et al. (2007) adopt LDRs to provide tractable formulations for solving ADRO problems. Bertsimas et al. (2018) incorporate the primary and auxiliary random variables of the lifted ambiguity set in the LDRs for ADRO problems, which significantly improves the solutions.

In this paper, we propose a high-level generic approach for ARO problems with fixed recourse via Fourier–Motzkin elimination (FME), which can be naturally integrated with existing approaches (e.g., decision rules, finite adaptability) to improve the quality of obtained solutions. FME was first introduced in Fourier (1826) and was rediscovered in Motzkin (1936). By using FME, we can reformulate the ARO problems into their equivalent counterparts with a reduced number of adjustable variables at the expense of an increasing number of constraints. From a theoretical standpoint, every ARO problem admits an equivalent static reformulation; however, one major obstacle in practice is that FME often leads to too many redundant constraints. To keep the resulting equivalent counterpart at its minimal size, after eliminating an adjustable variable via FME, we execute a linear programming (LP)-based procedure to detect and remove the redundant constraints. This redundant constraint identification (RCI) procedure is inspired by Caron et al. (1989). We propose to apply FME and RCI alternately to eliminate some of the adjustable variables and redundant constraints until the size of the reformulation reaches a prescribed computational limit, and then for the remaining adjustable variable, we impose LDRs to obtain an approximated solution. Zhen and den Hertog (2018) apply FME to compute the maximum volume inscribed ellipsoid of a polytopic projection.

Through the lens of FME, we investigate two-stage ARO problems theoretically and prove that there exist piecewise affine functions that are optimal decision rules (ODRs) for the adjustable variables. By applying FME to the dual formulation of Bertsimas and de Ruiter (2016), we further characterize the structures of the ODRs for a broad class of two-stage ARO problems: (a) we establish the optimality of LDRs for two-stage ARO problems with simplex uncertainty sets, and (b) for two-stage ARO problems with box uncertainty sets, we show that there exist two piecewise affine functions that are ODRs for the adjustable variables in the dual formulation, and these problems can be cast as sum-of-max problems. We further note that, despite the equivalence of primal and dual formulations, they may have significantly different numbers of adjustable variables. We evaluate the efficiency of our approach on both formulations numerically. By using our FME approach, we extend the approach of Bertsimas et al. (2018) for ADRO problems. Via numerical experiments, we show that our approach improves the obtained solutions in Bertsimas et al. (2018).

Our main contributions are as follows:

1. We present a high-level generic FME approach for ARO problems. We show that the FME approach is not necessarily a substitution for all existing methods, but can also be used before the existing methods are applied.

2. We investigate two-stage ARO problems via FME, which enables us to characterize the structures of the ODRs for a broad class of two-stage ARO problems.

3. We adapt an LP-based RCI procedure for ARO problems, which effectively removes the redundant constraints and improves the computability of the FME approach.

4. We show that our FME approach can be used to extend the approach of Bertsimas et al. (2018) for ADRO problems.

5. Using numerical experiments, we show that our approach can significantly improve the approximated solutions obtained from LDRs. Our approach is particularly effective for the formulations with few adjustable variables.

The rest of this paper is organized as follows. In Section 2, we introduce FME for two-stage ARO problems. Section 3 investigates the primal and dual formulations of two-stage ARO problems and presents some new results on the structures of the ODRs for several classes of two-stage ARO problems. In Section 4, we propose an LP-based RCI procedure to remove the redundant constraints. Section 5 uses our FME approach to extend the approach of Bertsimas et al. (2018) for ADRO problems. We generalize our approach to the multistage case in Section 6. Section 7 evaluates our approach numerically via lot-sizing on a network and appointment scheduling problems. Section 8 presents conclusions and future research.

Notations. We use $[N]$, $N \in \mathbb{N}$ to denote the set of running indices, $\{1, \ldots, N\}$. We generally use boldfaced characters such as $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{A} \in \mathbb{R}^{M \times N}$ to represent vectors and matrices, respectively, and $\mathbf{x}_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|}$ to denote a vector that contains a subset $\mathcal{S} \subseteq [N]$ of components in $\mathbf{x}$ (e.g., $x_i \in \mathbb{R}$ denotes the ith element of $\mathbf{x}$). We use $(x)^+$ and $|x|$ to denote $\max\{x, 0\}$ and the absolute value of $x \in \mathbb{R}$, and we use $|\mathcal{S}|$ to denote the cardinality of a finite set $\mathcal{S} \subseteq [N]$. Special vectors include $\mathbf{0}$, $\mathbf{1}$, and $\mathbf{e}_i$, which are, respectively, the vector of zeros, the vector of ones, and the standard unit basis vector. We denote $\mathcal{R}^{N,M}$ as the space of all measurable functions from $\mathbb{R}^N$ to $\mathbb{R}^M$ that are bounded on compact sets. We use a tilde to denote a random variable without associating it with a particular probability distribution. We use $\tilde{\mathbf{z}} \in \mathbb{R}^I$ to represent an I-dimensional random variable, and it can be associated with a probability distribution $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I)$, where $\mathcal{P}_0(\mathbb{R}^I)$ represents the set of all probability distributions on $\mathbb{R}^I$. We denote $\mathbb{E}_{\mathbb{P}}(\cdot)$ as the expectation with respect to the probability distribution $\mathbb{P}$. For a support set $\mathcal{W} \subseteq \mathbb{R}^I$, $\mathbb{P}(\tilde{\mathbf{z}} \in \mathcal{W})$ represents the probability of $\tilde{\mathbf{z}}$ being in $\mathcal{W}$ evaluated on the distribution $\mathbb{P}$.

2. Two-Stage Robust Optimization via Fourier–Motzkin Elimination

We first focus on a two-stage ARO problem where the first-stage, or *here-and-now*, decisions $\mathbf{x} \in \mathbb{R}^{N_1}$ are made before the realization of the uncertain parameters $\mathbf{z}$, and the second-stage, or *wait-and-see*, decisions $\mathbf{y}$ are made after the value of $\mathbf{z}$ is revealed, and $\mathbf{z}$ resides in a set $\mathcal{W} \subset \mathbb{R}^{I_1}$. Let us call $\mathbf{y}$ *adjustable variables*. With this setting, a two-stage ARO problem can be written as follows:

$$\min_{x \in \mathcal{X}} \mathbf{c}'\mathbf{x}, \tag{1}$$

where the *feasible set* $\mathcal{X}$ is the set of all feasible here-and-now decisions,

$$\mathcal{X} = \{\mathbf{x} \in X \mid \exists\, \mathbf{y} \in \mathcal{R}^{I_1, N_2}: \\ \mathbf{A}(\mathbf{z})\mathbf{x} + \mathbf{B}\mathbf{y}(\mathbf{z}) \geqslant \mathbf{d}(\mathbf{z}),\ \forall\, \mathbf{z} \in \mathcal{W}\}, \tag{2}$$

for a given domain $X \subseteq \mathbb{R}^{N_1}$ (e.g., $X = \mathbb{R}_+^{N_1}$ or $X = \mathbb{Z}^{N_1}$). Here, $\mathbf{A} \in \mathcal{R}^{I_1, M \times N_1}$, $\mathbf{d} \in \mathcal{R}^{I_1, M}$ are functions that map from the vector $\mathbf{z}$ to the input parameters of the linear optimization problem. Adopting the common assumptions in the robust optimization literature, these functions are affinely dependent on $\mathbf{z}$ and are given by

$$\mathbf{A}(\mathbf{z}) = \mathbf{A}^0 + \sum_{k \in [I_1]} \mathbf{A}^k z_k, \quad \mathbf{d}(\mathbf{z}) = \mathbf{d}^0 + \sum_{k \in [I_1]} \mathbf{d}^k z_k,$$

with $\mathbf{A}^0, \mathbf{A}^1, \ldots, \mathbf{A}^{I_1} \in \mathbb{R}^{M \times N_1}$ and $\mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{I_1} \in \mathbb{R}^M$. The matrix $\mathbf{B} \in \mathbb{R}^{M \times N_2}$, also known in stochastic programming as the *recourse matrix*, is constant, which corresponds to the stochastic programming format known as *fixed recourse*. For the case where the objective also includes the worst-case second-stage costs, it is well known that there is an equivalent epigraph reformulation that is in the form of problem (1). Although problem (1) may seem conservative as it does not exploit distributional knowledge of the uncertainties that may be available, Bertsimas et al. (2018) show that it is capable of modeling ADRO problems. In Section 5, we show how to apply our approach to solve ADRO problems. We then generalize our approach to the multistage case in Section 6. Problem (1) is generally intractable, even if there are only right-handside uncertainties (see Minoux 2011), because the adjustable variables $\mathbf{y}$ are decision rules instead of finite vectors of decision variables.

We propose to derive an equivalent representation of $\mathcal{X}$ by eliminating the adjustable variables $\mathbf{y}$ via FME. Algorithm 1 describes the FME procedure to eliminate an adjustable variable y_l, where $l \in [N_2]$. Here, we assume the feasible region of y_l is bounded for any $\mathbf{x} \in \mathcal{X}$. This algorithm is adapted from Bertsimas and Tsitsiklis (1997, p. 72) for polyhedral projections.

Note that the number of extra constraints after eliminating y_l equals $mn - m - n$, where $m = |\{i \mid b_{il} > 0\ \forall\, i \in [M]\}|$ and $n = |\{i \mid b_{il} < 0\ \forall\, i \in [M]\}|$, which can be determined before the elimination. Since Algorithm 1 does not affect the objective function or the uncertainty set $\mathcal{W}$ of Problem (1), Theorem 1 holds for ARO problems with general objective functions and uncertainty sets.

Theorem 1. $\mathcal{X} = \mathcal{X}_{\backslash\{l\}}$.

Algorithm 1 (Fourier–Motzkin Elimination for Two-Stage Problems)

1. For some $l \in [N_2]$, rewrite each constraint in $\mathcal{X}$ in the form: there exists $\mathbf{y} \in \mathcal{R}^{I_1, N_2}$ such that

$$b_{il} y_l(\mathbf{z}) \geqslant d_i(\mathbf{z}) - \sum_{j\in[N_1]} a_{ij}(\mathbf{z}) x_j - \sum_{j\in[N_2]\setminus\{l\}} b_{ij} y_j(\mathbf{z}), \quad \forall\, \mathbf{z} \in \mathcal{W},\ \forall\, i \in [M];$$

if $b_{il} \neq 0$, divide both sides by b_{il}. We obtain an equivalent representation of $\mathcal{X}$ involving the following constraints: there exists $\mathbf{y} \in \mathcal{R}^{I_1, N_2}$ such that

$$y_l(\mathbf{z}) \geqslant f_i(\mathbf{z}) + \mathbf{g}_i'(\mathbf{z})\mathbf{x} + \mathbf{h}_i'\mathbf{y}_{\setminus\{l\}}(\mathbf{z}), \quad \forall\, \mathbf{z} \in \mathcal{W} \quad \text{if } b_{il} > 0, \tag{3}$$

$$f_j(\mathbf{z}) + \mathbf{g}_j'(\mathbf{z})\mathbf{x} + \mathbf{h}_j'\mathbf{y}_{\setminus\{l\}}(\mathbf{z}) \geqslant y_l(\mathbf{z}), \quad \forall\, \mathbf{z} \in \mathcal{W} \quad \text{if } b_{jl} < 0, \tag{4}$$

$$0 \geqslant f_k(\mathbf{z}) + \mathbf{g}_k'(\mathbf{z})\mathbf{x} + \mathbf{h}_k'\mathbf{y}_{\setminus\{l\}}(\mathbf{z}), \quad \forall\, \mathbf{z} \in \mathcal{W} \quad \text{if } b_{kl} = 0. \tag{5}$$

Here, each $\mathbf{h}_i$, $\mathbf{h}_j$, $\mathbf{h}_k$ is a vector in $\mathbb{R}^{N_2-1}$ for a given $\mathbf{z}$; each f_i, f_j, f_k is a scalar; and each $\mathbf{g}_i$, $\mathbf{g}_j$, $\mathbf{g}_k$ is a vector in $\mathbb{R}^{N_1}$.

2. Let $\mathcal{X}_{\setminus\{l\}}$ be the feasible set after the adjustable variable y_l is eliminated, and it is defined by the following constraints: there exists $\mathbf{y}_{\setminus\{l\}} \in \mathcal{R}^{I_1, N_2-1}$ such that

$$f_j(\mathbf{z}) + \mathbf{g}_j'(\mathbf{z})\mathbf{x} + \mathbf{h}_j'\mathbf{y}_{\setminus\{l\}}(\mathbf{z}) \geqslant f_i(\mathbf{z}) + \mathbf{g}_i'(\mathbf{z})\mathbf{x} + \mathbf{h}_i'\mathbf{y}_{\setminus\{l\}}(\mathbf{z}), \quad \forall\, \mathbf{z} \in \mathcal{W} \quad \text{if } b_{jl} < 0 \text{ and } b_{il} > 0, \tag{6}$$

$$0 \geqslant f_k(\mathbf{z}) + \mathbf{g}_k'(\mathbf{z})\mathbf{x} + \mathbf{h}_k'\mathbf{y}_{\setminus\{l\}}(\mathbf{z}), \quad \forall\, \mathbf{z} \in \mathcal{W} \quad \text{if } b_{kl} = 0. \quad \square \tag{7}$$

Proof. This proof is adapted from Bertsimas and Tsitsiklis (1997, p. 73). If $\mathbf{x} \in \mathcal{X}$, there exist some vector functions $\mathbf{y}(\mathbf{z})$ such that $(\mathbf{x}, \mathbf{y}(\mathbf{z}))$ satisfies (3)–(5). It follows immediately that $(\mathbf{x}, \mathbf{y}_{\setminus\{l\}}(\mathbf{z}))$ satisfies (6) and (7), and $\mathbf{x} \in \mathcal{X}_{\setminus\{l\}}$. This shows that $\mathcal{X} \subset \mathcal{X}_{\setminus\{l\}}$.

We prove $\mathcal{X}_{\setminus\{l\}} \subset \mathcal{X}$. Let $\mathbf{x} \in \mathcal{X}_{\setminus\{l\}}$. It follows from (6) that there exists some $\mathbf{y}_{\setminus\{l\}}(\mathbf{z})$,

$$\min_{\{j \mid b_{jl}<0\}} \{f_j(\mathbf{z}) + \mathbf{g}_j'(\mathbf{z})\mathbf{x} + \mathbf{h}_j'\mathbf{y}_{\setminus\{l\}}(\mathbf{z})\} \geqslant \max_{\{i \mid b_{il}>0\}} \{f_i(\mathbf{z}) + \mathbf{g}_i'(\mathbf{z})\mathbf{x} + \mathbf{h}_i'\mathbf{y}_{\setminus\{l\}}(\mathbf{z})\}, \quad \forall\, \mathbf{z} \in \mathcal{W}.$$

Let

$$y_l(\mathbf{z}) = \theta \min_{\{j \mid b_{jl}<0\}} \{f_j(\mathbf{z}) + \mathbf{g}_j'(\mathbf{z})\mathbf{x} + \mathbf{h}_j'\mathbf{y}_{\setminus\{l\}}(\mathbf{z})\} + (1-\theta) \max_{\{i \mid b_{il}>0\}} \{f_i(\mathbf{z}) + \mathbf{g}_i'(\mathbf{z})\mathbf{x} + \mathbf{h}_i'\mathbf{y}_{\setminus\{l\}}(\mathbf{z})\}$$

for any $\theta \in [0,1]$. It then follows that $(\mathbf{x}, \mathbf{y}(\mathbf{z}))$ satisfies (3)–(5). Therefore, $\mathbf{x} \in \mathcal{X}$. $\square$

From Theorem 1, one can repeatedly apply Algorithm 1 to eliminate all the linear adjustable variables $\mathbf{y}$ in (2), which results in an equivalent set $\mathcal{X}_{\setminus[N_2]}$. The two-stage problem (1) can now be equivalently represented as a static robust optimization problem:

$$\min_{\mathbf{x}\in\mathcal{X}} \mathbf{c}'\mathbf{x} = \min_{\mathbf{x}\in\mathcal{X}_{\setminus[N_2]}} \mathbf{c}'\mathbf{x}. \tag{8}$$

If the uncertainty set $\mathcal{W}$ is convex, Problem (8) can be solved to optimality via the techniques from robust optimization (see, e.g., Mutapcic and Boyd 2009, Ben-Tal et al. 2015, Gorissen et al. 2014). However, in step 2 of Algorithm 1, the number of constraints may increase quadratically after each elimination. The complexity of eliminating N_2 adjustable variables from M constraints via Algorithm 1 is $\mathcal{O}(M^{2^{N_2}})$, which is an unfortunate inheritance of FME. In Section 4, we introduce an efficient LP-based procedure to detect and remove redundant constraints.

Example 1 (Lot Sizing on a Network). In lot sizing on a network, we have to determine the stock allocation x_i for $i \in [N]$ stores prior to knowing the realization of the demand at each location. The capacity of the stores is incorporated in X. The demand $\mathbf{z}$ is uncertain and assumed to be in an uncertainty set $\mathcal{W}$. After we observe the realization of the demand, we can transport stock y_{ij} from store i to store j at unit cost t_{ij} in order to meet all demand. The aim is to minimize the worst case storage costs (with unit costs c_i) and the cost arising from shifting the products from one store to another. The network flow model can now be written as a two-stage ARO problem:

$$\begin{aligned} \min_{\mathbf{x}\in X, y_{ij}, \tau} \quad & \mathbf{c}'\mathbf{x} + \tau \\ \text{s.t.} \quad & \sum_{i,j\in[N]} t_{ij} y_{ij}(\mathbf{z}) \leqslant \tau, && \forall \mathbf{z}\in\mathcal{W}, \\ & \sum_{j\in[N]} y_{ji}(\mathbf{z}) - \sum_{j\in[N]} y_{ij}(\mathbf{z}) \geqslant z_i - x_i, && \forall \mathbf{z}\in\mathcal{W},\ i\in[N], \\ & y_{ij}(\mathbf{z}) \geqslant 0, \quad y_{ij} \in \mathcal{R}^{N,1}, && \forall \mathbf{z}\in\mathcal{W},\ i,j\in[N]. \end{aligned} \tag{P}$$

The transportation cost $t_{ij} = 0$ if $i = j$ and $t_{ij} \geqslant 0$ if otherwise. For $N = 2$, there are four adjustable variables: y_{11}, y_{12}, y_{21}, and y_{22}. We apply Algorithm 1 iteratively, which leads to the following equivalent reformulation:

$$\begin{aligned} \min_{\mathbf{x}\in X, \tau} \quad & \mathbf{c}'\mathbf{x} + \tau \\ \text{s.t.} \quad & t_{21}z_1 - t_{21}x_1 \leqslant \tau, && \forall\, \mathbf{z} \in \mathcal{W}, \\ & t_{12}z_2 - t_{12}x_2 \leqslant \tau, && \forall\, \mathbf{z} \in \mathcal{W}, \\ & z_1 + z_2 - x_1 - x_2 \leqslant 0, && \forall\, \mathbf{z} \in \mathcal{W}, \\ & (t_{12} + t_{21})(z_1 - x_1) + t_{12}(x_2 - z_2) \leqslant \tau, && \forall\, \mathbf{z} \in \mathcal{W}. \end{aligned}$$

Note that we omit $\tau \geqslant 0$ because it is clearly a redundant constraint, which can be easily detected in the elimination procedure. This is a static robust linear optimization problem. We show in Section 7.1 that imposing linear decision rules on y_{ij} in P can lead to a suboptimal solution, whereas this equivalent reformulation produces the optimal solution. □

As a result of Algorithm 1, there may be many constraints in $\mathscr{X}_{\setminus[N_2]}$. We can first (iteratively) eliminate a subset $\mathscr{S} \subseteq [N_2]$ of the adjustable variables in $\mathscr{X}$ till the size of the resulting description $\mathscr{X}_{\setminus\mathscr{S}}$ reaches the prescribed computational limit, and then impose some simple functions (i.e., decision rules) $\mathscr{F}^{I_1,1} \subset \mathscr{R}^{I_1,1}$ on the remaining y_i, for all $i \in [N_2]\backslash\mathscr{S}$. The feasible set becomes

$$\hat{\mathscr{X}}_{\setminus\mathscr{S}} = \{\mathbf{x} \in X \mid \exists\, \mathbf{y}_{\setminus\mathscr{S}} \in \mathscr{F}^{I_1,N_2-|\mathscr{S}|}: \\ \mathbf{G}(\mathbf{z})\mathbf{x} + \mathbf{H}\mathbf{y}(\mathbf{z}) \geqslant \mathbf{f}(\mathbf{z}),\ \forall\, \mathbf{z} \in \mathscr{W}\},$$

where $\mathbf{G}(\mathbf{z})$ and $\mathbf{H}$ are the resulting coefficient matrices of $\mathbf{x}$ and $\mathbf{y}$, respectively, and $\mathbf{f}(\mathbf{z})$ is the corresponding right-hand-side vector after elimination. Since $\mathbf{y} \in \mathscr{F}^{I_1,N_2} \subset \mathscr{R}^{I_1,N_2}$, it follows that $\hat{\mathscr{X}}_{\setminus\mathscr{S}}$ is a conservative (inner) approximation of $\mathscr{X}_{\setminus\mathscr{S}}$ (i.e., $\hat{\mathscr{X}}_{\setminus\mathscr{S}} \subseteq \mathscr{X}_{\setminus\mathscr{S}}$). We simply use $\hat{\mathscr{X}}$ to denote $\hat{\mathscr{X}}_{\setminus\varnothing}$. The following theorem shows that the more adjustable variables are eliminated, the tighter the approximation becomes; if all the adjustable variables are eliminated, the set representation is exact (i.e., $\hat{\mathscr{X}}_{\setminus[N_2]} = \mathscr{X}_{\setminus[N_2]} = \mathscr{X}$).

Theorem 2. *$\hat{\mathscr{X}} \subseteq \hat{\mathscr{X}}_{\setminus\mathscr{S}_1} \subseteq \hat{\mathscr{X}}_{\setminus\mathscr{S}_2} \subseteq \mathscr{X}$, for all $\mathscr{S}_1 \subseteq \mathscr{S}_2 \subseteq [N_2]$.*

Proof. Let $\mathscr{S} \subseteq [N_2]$. After eliminating y_i, $i \in \mathscr{S}$, in $\mathscr{X}$ via Algorithm 1, we have

$$\mathscr{X}_{\setminus\mathscr{S}} = \{\mathbf{x} \in X \mid \exists\, \mathbf{y}_{\setminus\mathscr{S}} \in \mathscr{R}^{I_1,N_2-|\mathscr{S}|}: \\ \mathbf{G}(\mathbf{z})\mathbf{x} + \mathbf{H}\mathbf{y}_{\setminus\mathscr{S}}(\mathbf{z}) \geqslant \mathbf{f}(\mathbf{z}),\ \forall\, \mathbf{z} \in \mathscr{W}\},$$

where $\mathbf{G}(\mathbf{z})$ and $\mathbf{H}$ are the resulting coefficient matrices of $\mathbf{x}$ and $\mathbf{y}$, respectively, and $\mathbf{f}(\mathbf{z})$ is the corresponding right-hand-side vector after elimination. From Theorem 1, we know that $\mathscr{X}_{\setminus\mathscr{S}} = \mathscr{X}$. By imposing decision rules $\mathscr{F}^{I_1,1} \subset \mathscr{R}^{I_1,1}$ to the remaining y_i, for all $i \in [N_2]\backslash\mathscr{S}$, by definition, we have

$$\begin{aligned}\hat{\mathscr{X}}_{\setminus\mathscr{S}} &= \{\mathbf{x} \in X \mid \exists\, \mathbf{y}_{\setminus\mathscr{S}} \in \mathscr{F}^{I_1,N_2-|\mathscr{S}|}: \\ &\qquad \mathbf{G}(\mathbf{z})\mathbf{x} + \mathbf{H}\mathbf{y}_{\setminus\mathscr{S}}(\mathbf{z}) \geqslant \mathbf{f}(\mathbf{z}),\ \forall\, \mathbf{z} \in \mathscr{W}\} \\ &= \{\mathbf{x} \in X \mid \exists\, \mathbf{y}_{\setminus\mathscr{S}} \in \mathscr{F}^{I_1,N_2-|\mathscr{S}|}, \mathbf{y}_{\mathscr{S}} \in \mathscr{R}^{I_1,|\mathscr{S}|}: \\ &\qquad \mathbf{A}(\mathbf{z})\mathbf{x} + \mathbf{B}\mathbf{y}(\mathbf{z}) \geqslant \mathbf{d}(\mathbf{z}),\ \forall\, \mathbf{z} \in \mathscr{W}\},\end{aligned}$$

where $\mathbf{A}, \mathbf{B}$, and $\mathbf{d}$ are the same as in (2). Hence, it follows that $\hat{\mathscr{X}} \subseteq \hat{\mathscr{X}}_{\setminus\mathscr{S}} \subseteq \mathscr{X}_{\setminus\mathscr{S}} = \mathscr{X}$. Now, suppose that $\mathscr{S}_1 \subseteq \mathscr{S}_2 \subseteq [N_2]$; we have $\hat{\mathscr{X}}_{\setminus\mathscr{S}_1} \subseteq \hat{\mathscr{X}}_{\setminus\mathscr{S}_2} \subseteq \mathscr{X}_{\setminus\mathscr{S}_1} = \mathscr{X}_{\setminus\mathscr{S}_2} = \mathscr{X}$. □

Theorem 2 shows that Algorithm 1 can be used to improve the solutions of all existing methods, which include linear decision rules (see Ben-Tal et al. 2004, Chen and Zhang 2009), quadratic decision rules (see Ben-Tal et al. 2009), piecewise linear decision rules (see Chen et al. 2008, Chen and Zhang 2009, Bertsimas and Georghiou 2015), polynomial decision rules (see Bertsimas et al. 2011), and finite adaptability approaches. See Bertsimas and Dunning (2016), Postek and den Hertog (2016).

Algorithm 1 can also be applied to nonlinear ARO problems with a subset of adjustable variables appearing linearly in the constraints; for example, one can use Algorithm 1 to eliminate y_l in

$$\mathscr{X}^{ge} = \{\mathbf{x} \in X \mid \exists\, \mathbf{y} \in \mathscr{R}^{I_1,N_2}: \mathbf{f}(\mathbf{x}, \mathbf{y}_{\setminus\{l\}}, \mathbf{z}) + \mathbf{b}y_l \geqslant \mathbf{0},\ \forall\, \mathbf{z} \in \mathscr{W}\},$$

where $\mathbf{f} \in \mathscr{R}^{N_1\times(N_2-1)\times I_1,M}$ is a vector of general functions, and $\mathbf{b} \in \mathbb{R}^M$. Note that the constraints in $\mathscr{X}^{ge}$ are convex or concave in $\mathbf{x}$, $\mathbf{y}$, and/or $\mathbf{z}$; the constraints in $\mathscr{X}^{ge}_{\setminus\{l\}}$ remain convex or concave in $\mathbf{x}$, $\mathbf{y}_{\setminus\{l\}}$, and/or $\mathbf{z}$, $l \in [N_2]$.

It is worth noting that, for ARO problems without (*relatively*) *complete recourse*, imposing simple decision rules may lead to infeasibility. For those ARO problems, one can first eliminate some of the adjustable variables to "enlarge" the feasible region (see Theorem 2), then solve them via decision rules or finite adaptability approaches. We emphasize that our approach is not necessarily a substitution of all existing methods, but it can also be used before the existing methods are applied as a kind of preprocessing. For the rest of this paper, we mainly focus on the two-stage robust linear optimization model (1) and illustrate the effectiveness of our approach by complementing the most conventional method—that is, LDR, $\mathbf{y} \in \mathscr{F}^{I_1,N_2} = \mathscr{L}^{I_1,N_2}$, where

$$\mathscr{L}^{I_1,N_2} = \left\{\mathbf{y} \in \mathscr{R}^{I_1,N_2} \;\middle|\; \begin{array}{l} \exists\, \mathbf{y}^0, \mathbf{y}^i \in \mathbb{R}^{N_2}, \quad i \in [I_1]: \\ \mathbf{y}(\mathbf{z}) = \mathbf{y}^0 + \sum_{i\in[I_1]} \mathbf{y}^i z_i \end{array}\right\},$$

and $\mathbf{y}^i \in \mathbb{R}^{N_2}$, $i \in [I_1] \cup \{0\}$, are decision variables. We show that for small-sized ARO problems, our approach gives a static tractable counterpart of the ARO problems and finds the optimal solution. For moderate- or large-sized instances, we eliminate a subset of the adjustable variables and then impose LDR on the remaining adjustable variables. This yields provably better solutions than imposing LDR on all adjustable variables.

3. Optimality of Decision Rules: A Primal–Dual Perspective

In this section, we investigate the primal and dual formulations of two-stage ARO problems through the lens of FME, which enables us to derive some new results on the optimality of certain decision rule structures for several classes of problems.

3.1. A Primal Perspective

As an immediate consequence of Algorithm 1, one can prove the following result for two-stage ARO problems.

Theorem 3. *There exist ODRs for problem (1) such that y_l, $l \in [N_2]$, is a convex piecewise affine function or a concave piecewise affine function, and the remaining components of $\mathbf{y}$ are general piecewise affine functions.*

Proof. Let us denote $\mathbf{x}^*$ as the optimal here-and-now decisions and eliminate all but one adjustable variable y_l in $\mathcal{X}$ defined in (2) via Algorithm 1. Let $\mathcal{S}_l = [N_2]\backslash\{l\}$. From Theorem 1, we know $\mathcal{X} = \mathcal{X}_{\backslash \mathcal{S}_l}$. The adjustable variable y_l is upper (lower) bounded by a finite number of minimum (maximum) of affine functions in $\mathbf{z}$; that is,

$$\check{f}_l(\mathbf{z}) \leqslant y_l(\mathbf{z}) \leqslant \hat{f}_l(\mathbf{z}), \quad \forall\, \mathbf{z} \in \mathcal{W}, \tag{9}$$

where $\check{f}_l(\mathbf{z})$ and $\hat{f}_l(\mathbf{z})$ are, respectively, convex piecewise affine and concave piecewise affine functions of $\mathbf{z} \in \mathcal{W}$. If problem (1) is feasible, then the constraint

$$\check{f}_l(\mathbf{z}) \leqslant \hat{f}_l(\mathbf{z}), \quad \forall\, \mathbf{z} \in \mathcal{W}$$

must hold, and hence $y_l(\mathbf{z}) = \check{f}_l(\mathbf{z})$ and $y_l(\mathbf{z}) = \hat{f}_l(\mathbf{z})$ would be ODRs for the adjustable variable y_l in problem (1) for all $l \in [N_2]$. Once the ODR of y_l is determined, one can then determine the ODR of the last eliminated adjustable variable from its upper and lower bounding functions as in (9). The ODR of the second-to-last eliminated adjustable variable can be determined analogously. The ODRs of the adjustable variables can be determined iteratively by reversing Algorithm 1 in the exact reversed order of the eliminations. It follows that there exist piecewise affine functions (not necessarily concave or convex) that are ODRs for the adjustable variables y_i, $i \in [N_2]\backslash\{l\}$ in problem (1). □

Since we do not impose any assumption on the uncertainty set $\mathcal{W}$ in problem (1), Theorem 3 holds for problem (1) with general uncertainty sets. Bemporad et al. (2003) show a similar result for problem (1) with right-hand-side polyhedral uncertainties. Motivated by the result of Bemporad et al. (2003), in Bertsimas and Georghiou (2015) and Ben-Tal et al. (2016), the authors construct piecewise linear decision rules for ARO problems with right-hand-side polyhedral uncertainties. Theorem 3 stimulates a generalization of the existing methods for ARO problems with uncertainties that (a) reside in general convex sets or (b) appear on both sides of the constraints.

3.2. A Dual Perspective for Polyhedral Uncertainty Sets

Given a polyhedral uncertainty set

$$\mathcal{W}_{\text{poly}} = \{\mathbf{z} \in \mathbb{R}^{I_1} \mid \exists\, \mathbf{v} \in \mathbb{R}^{I_2} \colon \mathbf{P}'\mathbf{z} + \mathbf{Q}'\mathbf{v} \leqslant \boldsymbol{\rho}\},$$

where $\mathbf{P} \in \mathbb{R}^{I_1 \times K}$, $\mathbf{Q} \in \mathbb{R}^{I_2 \times K}$ and $\boldsymbol{\rho} \in \mathbb{R}^K$, Bertsimas and de Ruiter (2016) derive an equivalent dual formulation of problem (1) (see the proof in the appendix):

$$\min_{\mathbf{x} \in \mathcal{X}^D} \mathbf{c}'\mathbf{x}, \tag{10}$$

where the equivalent *dual feasible set* $\mathcal{X}^D$ (i.e., $\mathcal{X}^D = \mathcal{X}$) is defined as follows:

$$\mathcal{X}^D = \left\{ \mathbf{x} \in X \;\middle|\; \exists\, \boldsymbol{\lambda} \in \mathcal{R}^{M,K} \colon \begin{array}{r} \boldsymbol{\omega}'(\mathbf{A}^0\mathbf{x} - \mathbf{d}^0) - \boldsymbol{\rho}'\boldsymbol{\lambda}(\boldsymbol{\omega}) \geqslant \mathbf{0}, \\ \forall\, \boldsymbol{\omega} \in \mathcal{U} \\ \mathbf{p}_i'\boldsymbol{\lambda}(\boldsymbol{\omega}) = (\mathbf{d}^i - \mathbf{A}^i\mathbf{x})'\boldsymbol{\omega}, \\ \forall\, \boldsymbol{\omega} \in \mathcal{U},\ \forall\, i \in [I_1] \\ \mathbf{Q}\boldsymbol{\lambda}(\boldsymbol{\omega}) = \mathbf{0}, \quad \boldsymbol{\lambda}(\boldsymbol{\omega}) \geqslant \mathbf{0}, \\ \forall\, \boldsymbol{\omega} \in \mathcal{U} \end{array} \right\}, \tag{11}$$

with the *dual uncertainty set*:

$$\mathcal{U} = \{\boldsymbol{\omega} \in \mathbb{R}_+^M \mid \mathbf{B}'\boldsymbol{\omega} = \mathbf{0}\},$$

where $\mathbf{p}_i \in \mathbb{R}^{I_1}$ are the ith row vectors of matrix $\mathbf{P}$ for $i \in [I_1]$. There exist auxiliary variables $\mathbf{v}$ in $\mathcal{W}_{\text{poly}}$. For the decision rules of problem (1), the adjustable variables $\mathbf{y}$ should depend on both $\mathbf{z}$ and $\mathbf{v}$. Bertsimas and de Ruiter (2016) show that primal and dual formulations with LDRs are also equivalent, and optimal LDRs for one formulation can be easily constructed from the solution of the other formulation by solving a system of linear equations. The equalities in (11) can be used to eliminate some of the adjustable variables $\boldsymbol{\lambda}$ via Gaussian elimination. Zhen and den Hertog (2018) show that eliminating adjustable variables in the equalities of a two-stage ARO problem is equivalent to imposing LDRs.

One can apply Algorithm 1 to eliminate adjustable variables in the dual formulation (10). Note that the structure of the uncertainty set in the primal formulation (1) becomes part of the constraints in the dual formulation (10). From Theorem 3, there exist piecewise affine functions that are ODRs for the adjustable variables $\boldsymbol{\lambda}$ in the dual formulation (10). Let us consider two special classes of $\mathcal{W}_{\text{poly}}$: a standard simplex and a box.

Theorem 4. *Suppose that the uncertainty set $\mathcal{W}_{\text{poly}}$ is a standard simplex. Then, there exist LDRs that are ODRs for the adjustable variables $\mathbf{y}$ in problem (1).*

Proof. Suppose that $\mathbf{z}$ reside in a standard simplex:

$$\mathcal{W}_{\text{simplex}} = \{\mathbf{z} \in \mathbb{R}_+^{I_1} \mid \mathbf{1}'\mathbf{z} \leqslant 1\}.$$

From (11), we have the following reformulation:

$$\mathcal{X}^D = \left\{ \mathbf{x} \in X \;\middle|\; \exists\, \lambda \in \mathcal{R}^{M,1} \colon \begin{array}{r} \boldsymbol{\omega}'(\mathbf{A}^0\mathbf{x} - \mathbf{d}^0) - \lambda(\boldsymbol{\omega}) \geqslant \mathbf{0}, \\ \forall\, \boldsymbol{\omega} \in \mathcal{U} \\ \lambda(\boldsymbol{\omega}) \geqslant ((\mathbf{d}^i - \mathbf{A}^i\mathbf{x})'\boldsymbol{\omega})^+, \\ \forall\, \boldsymbol{\omega} \in \mathcal{U},\ \forall\, i \in [I_1] \end{array} \right\}.$$

Observe that the dual adjustable variable $\lambda(\mathbf{w})$ is feasible in $\mathscr{X}^D$ if and only if

$$\boldsymbol{\omega}'(\mathbf{A}^0\mathbf{x}-\mathbf{d}^0) \geqslant \lambda(\boldsymbol{\omega}) \geqslant \left(\max_{i\in[I_1]}\{(\mathbf{d}^i-\mathbf{A}^i\mathbf{x})'\boldsymbol{\omega}\}\right)^+, \quad \forall\, \boldsymbol{\omega}\in\mathscr{U}.$$

Hence, there exists an ODR in the form of $\lambda(\boldsymbol{\omega}) = \boldsymbol{\omega}'(\mathbf{A}^0\mathbf{x}-\mathbf{d}^0)$, which is affine in $\boldsymbol{\omega}$. Using the techniques of Bertsimas and de Ruiter (2016, theorem 2), we can construct optimal LDRs for the adjustable variables $\mathbf{y}$ in the primal formulation (1). □

Theorem 4 coincides with the recent finding in Ben-Ameur et al. (2017, corollary 2), which is a generalization of the result of Bertsimas and Goyal (2012, theorem 1) where authors prove there exist LDRs that are optimal for two-stage ARO problems with only right-hand-side uncertainties that reside in a simplex set. Zhen and den Hertog (2018) use Theorem 4 to prove that there exist polynomials of (at most) degree I_1 and linear in each z_i, $\forall\, i\in[I_1]$, that are ODRs for $\mathbf{y}$ in problem (1) with general convex uncertainty sets.

Theorem 5. *Suppose that the uncertainty set $\mathscr{W}_{\text{poly}}$ is a box. Then, the convex two-piecewise affine functions in the form of $((\mathbf{d}^i-\mathbf{A}^i\mathbf{x})'\boldsymbol{\omega})^+$ are ODRs for the adjustable variables λ_i, $i\in[I_1]$ in problem (10).*

Proof. Suppose $\mathbf{z}$ resides in the box:

$$\mathscr{W}_{box} = \{\mathbf{z}\in\mathbb{R}^{I_1} \mid -\boldsymbol{\rho}\leqslant\mathbf{z}\leqslant\boldsymbol{\rho}\},$$

where $\boldsymbol{\rho}\in\mathbb{R}^{I_1}_+$. After eliminating the equalities in (11) via Gaussian elimination, we have the following reformulation:

$$\mathscr{X}^D = \left\{ \mathbf{x}\in X \;\middle|\; \exists\boldsymbol{\lambda}\in\mathscr{R}^{I_1,I_1}\colon \begin{array}{c} \boldsymbol{\omega}'(\mathbf{A}^0\mathbf{x}-\mathbf{d}^0)+\sum_{i\in[I_1]}\rho_i(\mathbf{A}^i\mathbf{x}-\mathbf{d}^i)'\boldsymbol{\omega} \\ -2\boldsymbol{\rho}'\boldsymbol{\lambda}(\boldsymbol{\omega})\geqslant\mathbf{0}, \quad \forall\,\boldsymbol{\omega}\in\mathscr{U} \\ \lambda_i(\boldsymbol{\omega})\geqslant((\mathbf{A}^i\mathbf{x}-\mathbf{d}^i)'\boldsymbol{\omega})^+, \\ \forall\,\boldsymbol{\omega}\in\mathscr{U}, \quad \forall\, i\in[I_1] \end{array} \right\}. \tag{12}$$

After eliminating all but one adjustable variable λ_l in $\mathscr{X}^D$ via Algorithm 1, $l\in[I_1]$, the dual adjustable variable $\lambda_l(\boldsymbol{\omega})$ is feasible in $\mathscr{X}^D$ if and only if

$$\begin{aligned} &\frac{1}{2\rho_l}\Bigg[\boldsymbol{\omega}'(\mathbf{A}^0\mathbf{x}-\mathbf{d}^0)+\sum_{i\in[I_1]}\rho_i(\mathbf{A}^i\mathbf{x}-\mathbf{d}^i)'\boldsymbol{\omega} \\ &\qquad -\sum_{i\in[I_1]\setminus\{l\}}2\rho_i((\mathbf{A}^i\mathbf{x}-\mathbf{d}^i)'\boldsymbol{\omega})^+\Bigg] \geqslant \lambda_l(\boldsymbol{\omega}), \quad \forall\,\boldsymbol{\omega}\in\mathscr{U}; \\ &((\mathbf{A}^i\mathbf{x}-\mathbf{d}^i)'\boldsymbol{\omega})^+ \leqslant \lambda_l(\boldsymbol{\omega}), \qquad \forall\,\boldsymbol{\omega}\in\mathscr{U}. \end{aligned}$$

One can observe that λ_l is upper bounded by a $2^{|I_1-1|}$-piecewise affine function and lower bounded by $((\mathbf{d}^l-\mathbf{A}^l\mathbf{x})'\boldsymbol{\omega})^+$. Hence, there exists an ODR in the form of $\lambda_l(\boldsymbol{\omega})=((\mathbf{d}^l-\mathbf{A}^l\mathbf{x})'\boldsymbol{\omega})^+$—that is, a two-piecewise affine function. Analogously, it follows that there exist ODRs in the form of $\lambda_i(\boldsymbol{\omega})=((\mathbf{d}^i-\mathbf{A}^i\mathbf{x})'\boldsymbol{\omega})^+$ for all $i\in[I_1]$. □

An immediate observation from Theorem 5 is that, if we eliminate all the adjustable variables in (12) via Algorithm 1, it results in a sum-of-max representation:

$$\mathscr{X}^D_{\setminus[I_1]} = \left\{\mathbf{x}\in X \;\middle|\; \forall\,\boldsymbol{\omega}\in\mathscr{U}\colon \boldsymbol{\omega}'(\mathbf{A}^0\mathbf{x}-\mathbf{d}^0) \geqslant \sum_{i\in[I_1]}\rho_i|(\mathbf{d}^i-\mathbf{A}^i\mathbf{x})'\boldsymbol{\omega}|\right\}.$$

Note that there is only one constraint. One can use the techniques proposed in Gorissen and den Hertog (2013) and Ardestani-Jaafari and Delage (2016b) to solve problem (1) with box uncertainties approximately.

Note that the number of the uncertain parameters $\boldsymbol{\omega}\in\mathscr{U}$ in the dual formulation (10) equals the number of constraints in the primal formulation (1). Therefore, reducing the number of adjustable variables in the primal (via Algorithm 1), which leads to more constraints, is equivalent to lifting the uncertainty set of the dual formulation into higher dimensions. In other words, Algorithm 1 can also be interpreted as a lifting operation that lifts the polyhedral uncertainty sets of ARO problems into higher dimensions to enhance the decision rules. A related method is proposed by Chen and Zhang (2009), where the authors improve LDR-based approximations for ARO problems with fixed recourse by lifting the norm-based uncertainty sets into higher dimensions.

One could also use FME sequentially for the primal and dual formulation. Step 1. eliminate (a subset of) the adjustable variables $\mathbf{y}$ in the primal (1); Step 2. derive the corresponding dual formulation; Step 3. eliminate some adjustable variables in the obtained dual formulation; Step 4. solve the resulting problem via decision rules (if not all of the adjustable variables are eliminated). Since in Section 7.1 we will see that the dual formulation is far more effective than the primal, we do not consider this sequential procedure in our numerical experiments.

4. Redundant Constraint Identification

It is well known that Fourier–Motzkin elimination often leads to many redundant constraints. In this section, we present a simple yet effective LP-based procedure to remove those redundant constraints. First, we give a formal definition of redundant constraints for ARO problems.

Definition 1. We say that the lth constraint, $l\in[M]$, in the feasible set (2) is redundant if and only if for all $\mathbf{x}\in X$ and $\mathbf{y}\in\mathscr{R}^{I_1,N_2}$ such that

$$\mathbf{a}_i'(\boldsymbol{\zeta})\mathbf{x}+\mathbf{b}_i'\mathbf{y}(\boldsymbol{\zeta})\geqslant d_i(\boldsymbol{\zeta}), \quad \forall\,\boldsymbol{\zeta}\in\mathscr{W},\ \forall\, i\in[M]\setminus\{l\} \tag{13}$$

then

$$\mathbf{a}_l'(\mathbf{z})\mathbf{x}+\mathbf{b}_l'\mathbf{y}(\mathbf{z})\geqslant d_l(\mathbf{z}), \quad \forall\,\mathbf{z}\in\mathscr{W}, \tag{14}$$

where $\mathbf{a}_i$ and $\mathbf{b}_i$ are the ith row vectors of matrices $\mathbf{A}$ and $\mathbf{B}$, respectively, and d_i is the ith component of $\mathbf{d}$ for $i \in [M]$.

Hence, a redundant constraint is implied by the other constraints in (2), and it does not define the feasible region of $\mathbf{x}$. The RCI procedure in Theorem 6 is inspired by Caron et al. (1989).

Theorem 6. *The lth constraint, $l \in [M]$ in the feasible set (2) is redundant if and only if*

$$\begin{aligned} Z_l^* = \min_{\mathbf{x},\mathbf{y},\mathbf{z}}\ & \mathbf{a}_l'(\mathbf{z})\mathbf{x}+\mathbf{b}_l'\mathbf{y}(\mathbf{z})-d_l(\mathbf{z})\\ \text{s.t.}\ & \mathbf{a}_i'(\boldsymbol{\zeta})\mathbf{x}+\mathbf{b}_i'\mathbf{y}(\boldsymbol{\zeta})\geqslant d_i(\boldsymbol{\zeta}),\quad \forall \boldsymbol{\zeta}\in\mathcal{W},\ \forall i\in[M]\backslash\{l\},\\ & \mathbf{x}\in X,\quad \mathbf{y}\in\mathcal{R}^{I_1,N_2},\quad \mathbf{z}\in\mathcal{W}, \end{aligned} \tag{15}$$

has a nonnegative optimal objective (i.e., $Z_l^* \geqslant 0$).

Proof. Indeed, if $Z_l^* \geqslant 0$, then for all $\mathbf{x} \in X$ and $\mathbf{y} \in \mathcal{R}^{I_1,N_2}$ that are feasible in (13), we also have

$$0 \leqslant Z_l^* \leqslant \min_{\mathbf{z}\in\mathcal{W}}\{\mathbf{a}_l'(\mathbf{z})\mathbf{x}+\mathbf{b}_l'\mathbf{y}(\mathbf{z})-d_l(\mathbf{z})\},$$

which implies feasibility in (14). Conversely, if $Z_l^* < 0$, from the optimum solution of problem (15), there exists a solution $\mathbf{x} \in X$ and $\mathbf{y} \in \mathcal{R}^{I_1,N_2}$ that would be feasible in (13), but

$$\min_{\mathbf{z}\in\mathcal{W}}\{\mathbf{a}_l'(\mathbf{z})\mathbf{x}+\mathbf{b}_l'\mathbf{y}(\mathbf{z})-d_l(\mathbf{z})\} < 0,$$

which would be infeasible in (14). □

Unfortunately, identifying a redundant constraint could be as hard as solving the ARO problem. Moreover, not all redundant constraints have to be eliminated, since only the constraints with adjustable variables are potentially "malignant" and could lead to proliferations of redundant constraints after Algorithm 1. Therefore, we propose the following heuristic for identifying a potential malignant redundant constraint (i.e, one that has adjustable variables).

Theorem 7. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two disjoint subsets of $[M]$ such that*

$$\begin{aligned} \mathbf{a}_i(\mathbf{z}) = \mathbf{a}_i,\ \ \mathbf{b}_i \neq \mathbf{0},&\quad \forall i \in \mathcal{M}_1,\\ \mathbf{b}_i = \mathbf{0},&\quad \forall i \in \mathcal{M}_2. \end{aligned}$$

Then the lth constraint $l \in \mathcal{M}_1$ in the feasible set (2) is redundant if the following tractable static RO problem,

$$\begin{aligned} Z_l^\dagger = \min_{\mathbf{x},\mathbf{y},\mathbf{z}}\ & \mathbf{a}_l'\mathbf{x}+\mathbf{b}_l'\mathbf{y}-d_l(\mathbf{z})\\ \text{s.t.}\ & \mathbf{a}_i'(\boldsymbol{\zeta})\mathbf{x}\geqslant d_i(\boldsymbol{\zeta}),\quad \forall \boldsymbol{\zeta}\in\mathcal{W},\ \forall i\in\mathcal{M}_2,\\ & \mathbf{a}_i'\mathbf{x}+\mathbf{b}_i'\mathbf{y}\geqslant d_i(\mathbf{z}),\quad \forall i\in\mathcal{M}_1\backslash\{l\},\\ & \mathbf{x}\in X,\quad \mathbf{y}\in\mathbb{R}^{N_2},\quad \mathbf{z}\in\mathcal{W}, \end{aligned} \tag{16}$$

has a nonnegative optimal objective value (i.e., $Z_l^\dagger \geqslant 0$).

Proof. Observe that for any $l \in \mathcal{M}_1$,

$$\begin{aligned} Z_l^* \geqslant \min_{\mathbf{x},\mathbf{y},\mathbf{z}}\ & \mathbf{a}_l'\mathbf{x}+\mathbf{b}_l'\mathbf{y}(\mathbf{z})-d_l(\mathbf{z})\\ \text{s.t.}\ & \mathbf{a}_i'(\boldsymbol{\zeta})\mathbf{x}\geqslant d_i(\boldsymbol{\zeta}),\quad \forall \boldsymbol{\zeta}\in\mathcal{W},\ \forall i\in\mathcal{M}_2,\\ & \mathbf{a}_i'\mathbf{x}+\mathbf{b}_i'\mathbf{y}(\boldsymbol{\zeta})\geqslant d_i(\boldsymbol{\zeta}):\\ & \qquad\qquad \forall \boldsymbol{\zeta}\in\mathcal{W},\ \forall i\in\mathcal{M}_1\backslash\{l\},\\ & \mathbf{x}\in X,\quad \mathbf{y}\in\mathcal{R}^{I_1,N_2},\quad \mathbf{z}\in\mathcal{W},\\ Z_l^* \geqslant \min_{\mathbf{x},\mathbf{y},\mathbf{z}}\ & \mathbf{a}_l'\mathbf{x}+\mathbf{b}_l'\mathbf{y}(\mathbf{z})-d_l(\mathbf{z})\\ \text{s.t.}\ & \mathbf{a}_i'(\boldsymbol{\zeta})\mathbf{x}\geqslant d_i(\boldsymbol{\zeta}),\quad \forall \boldsymbol{\zeta}\in\mathcal{W},\ \forall i\in\mathcal{M}_2,\\ & \mathbf{a}_i'\mathbf{x}+\mathbf{b}_i'\mathbf{y}(\mathbf{z})\geqslant d_i(\mathbf{z}),\quad \forall i\in\mathcal{M}_1\backslash\{l\},\\ & \mathbf{x}\in X,\quad \mathbf{y}\in\mathcal{R}^{I_1,N_2},\quad \mathbf{z}\in\mathcal{W},\\ Z_l^* = \min_{\mathbf{x},\mathbf{y},\mathbf{z}}\ & \mathbf{a}_l'\mathbf{x}+\mathbf{b}_l'\mathbf{y}-d_l(\mathbf{z})\\ \text{s.t.}\ & \mathbf{a}_i'(\boldsymbol{\zeta})\mathbf{x}\geqslant d_i(\boldsymbol{\zeta}),\quad \forall \boldsymbol{\zeta}\in\mathcal{W},\ \forall i\in\mathcal{M}_2,\\ & \mathbf{a}_i'\mathbf{x}+\mathbf{b}_i'\mathbf{y}\geqslant d_i(\mathbf{z}),\quad \forall i\in\mathcal{M}_1\backslash\{l\},\\ & \mathbf{x}\in X,\quad \mathbf{y}\in\mathbb{R}^{N_2},\quad \mathbf{z}\in\mathcal{W},\\ Z_l^* = Z_l^\dagger. & \end{aligned}$$

Hence, whenever $Z_l^\dagger \geqslant 0$, we have $Z_l^* \geqslant 0$, implying that the lth constraint is redundant. □

Note that in Theorem 7, to avoid intractability, only a subset of constraints in the feasible set (2) is considered (i.e., $\mathcal{M}_1 \cup \mathcal{M}_2 \neq [M]$). We can extend the subset $\mathcal{M}_1 \subseteq [M]$ if the uncertainties affecting the constraints in $\mathcal{M}_1$ are columnwise. Specifically, let $\{\mathbf{z}^0, \ldots, \mathbf{z}^{N_1}\}$, $\mathbf{z}^j \in \mathbb{R}^{I_1^j}$, $j \in [N_1] \cup \{0\}$ be a partition of the vector $\mathbf{z} \in \mathbb{R}^{I_1}$ into $N_1 + 1$ vectors (including empty ones) such that

$$\mathcal{W} = \{(\mathbf{z}^0, \ldots, \mathbf{z}^{N_1}) \mid \mathbf{z}^j \in \mathcal{W}_j,\ \forall j \in [N_1] \cup \{0\}\}. \tag{17}$$

Note that if $\mathbf{z}^j$, $j \in [N_1]$ are empty vectors, then we would have $\mathcal{W} = \mathcal{W}_0$. Let $\mathcal{S} \subseteq [N_1]$ and $\bar{\mathcal{S}} = [N_1]\backslash\mathcal{S}$ such that $x_j \geqslant 0$ for all $j \in \mathcal{S}$ is implied by the set X. We redefine the subset $\mathcal{M}_1 \subseteq [M]$ such that for all $i \in \mathcal{M}_1$, $\mathbf{b}_i \neq \mathbf{0}$ and the functions $a_{ij} \in \mathcal{L}^{I_1^j,1}$ and $d_i \in \mathcal{L}^{I_1^0,1}$ are affine in $\mathbf{z}^j$ for all $j \in [N_2] \cup \{0\}$; specifically,

$$\begin{aligned} a_{ij}(\mathbf{z}) = a_{ij}(\mathbf{z}^j),&\quad \forall j \in \mathcal{S},\\ a_{ij}(\mathbf{z}) = a_{ij},&\quad \forall j \in \bar{\mathcal{S}},\\ d_i(\mathbf{z}) = d_i(\mathbf{z}^0).& \end{aligned}$$

Note that since $\mathcal{S}$ or $\mathbf{z}^j$, $j \in [N_1]$ can be empty sets, the conditions to select $\mathcal{M}_1$ are more general than in Theorem 7. From Theorem 7, one can check whether the lth inequality, $l \in \mathcal{M}_1$, is redundant by solving the following problem:

$$\begin{aligned} Z_l^\dagger = \min_{\mathbf{x}\in X,\mathbf{y},\mathbf{z}}\ & \sum_{j\in\mathcal{S}} a_{lj}(\mathbf{z}^j)x_j + \sum_{j\in\bar{\mathcal{S}}} a_{lj}x_j + \mathbf{b}_l'\mathbf{y} - d_l(\mathbf{z}^0)\\ \text{s.t.}\ & \mathbf{a}_i'(\boldsymbol{\zeta})\mathbf{x}\geqslant d_i(\boldsymbol{\zeta}),\quad \forall \boldsymbol{\zeta}\in\mathcal{W},\ \forall i\in\mathcal{M}_2,\\ & \sum_{j\in\mathcal{S}} a_{ij}(\mathbf{z}^j)x_j + \sum_{j\in\bar{\mathcal{S}}} a_{ij}x_j + \mathbf{b}_i'\mathbf{y} \geqslant d_i(\mathbf{z}^0),\\ & \qquad\qquad\qquad \forall i\in\mathcal{M}_1\backslash\{l\},\\ & \mathbf{z}^j\in\mathcal{W}_j,\qquad \forall j\in\mathcal{S}\cup\{0\}. \end{aligned} \tag{18}$$

The lth inequality is redundant if the optimal objective value is nonnegative. Because of the presence of products of variables (e.g., $\mathbf{z}^j x_j$), problem (18) is nonconvex in $\mathbf{x}$ and $\mathbf{z}$. An equivalent convex representation of (18) can be obtained by substituting $\mathbf{w}^j = \mathbf{z}^j x_j$, $j \in \mathcal{S}$:

$$
\begin{aligned}
Z_l^{\ddagger} = \min_{\mathbf{x}\in X,\mathbf{y},\mathbf{z}} \ & \sum_{j\in\mathcal{S}} a_{lj}(\mathbf{w}^j/x_j)x_j + \sum_{j\in\bar{\mathcal{S}}} a_{lj}x_j + \mathbf{b}_l'\mathbf{y} - d_l(\mathbf{z}^0)\\
\text{s.t.}\ & \mathbf{a}_i'(\boldsymbol{\zeta})\mathbf{x} \geqslant d_i(\boldsymbol{\zeta}), && \forall \boldsymbol{\zeta}\in\mathcal{W},\ \forall i \in \mathcal{M}_2,\\
& \sum_{j\in\mathcal{S}} a_{ij}(\mathbf{w}^j/x_j)x_j + \sum_{j\in\bar{\mathcal{S}}} a_{ij}x_j + \mathbf{b}_i'\mathbf{y} \geqslant d_i(\mathbf{z}^0), \\
& && \forall i \in \mathcal{M}_1 \setminus \{l\},\\
& (\mathbf{w}^j, x_j) \in \mathcal{K}_j, && \forall j \in \mathcal{S},\\
& \mathbf{z}^0 \in \mathcal{W}_0,
\end{aligned}
\tag{19}
$$

where $a_{ij}(\mathbf{w}^j/x_j)x_j$ is linear in $(\mathbf{w}^j, x_j)$, and the set $\mathcal{K}_j$ is a convex cone defined as

$$
\mathcal{K}_j = \mathrm{cl}\{(\mathbf{u}, t) \in \mathbb{R}^{I_1^j+1} \mid \mathbf{u}/t \in \mathcal{W}_j,\ t > 0\}.
$$

Hence, (19) is a convex optimization problem. This transformation technique is first proposed in Dantzig (1963) to solve generalized linear programs. Gorissen et al. (2014) use this technique to derive tractable robust counterparts of a linear conic optimization problem. Zhen and den Hertog (2017) apply this technique to derive a convex representation of the feasible set for systems of uncertain linear equations.

Algorithm 1 does not destroy the columnwise uncertainties, and the resulting reformulations from Algorithm 1 and RCI procedure are independent from the objective function of ARO problems. Therefore, the reformulation can be precomputed off-line and used to evaluate different objectives. Two-stage ARO problems with columnwise uncertainties are considered in, for example, Minoux (2011), Ardestani-Jaafari and Delage (2016a), and Xu and Burer (2018).

Example 2 (Removing Redundant Constraints for Lot Sizing on a Network). Let us again consider P in Example 1. The uncertain demand $\mathbf{z}$ is assumed to be in a budget uncertainty set:

$$
\mathcal{W} = \{z \in \mathbb{R}_+^N \mid \mathbf{z} \leqslant \mathbf{20}, \mathbf{1}'\mathbf{z} \leqslant 20\sqrt{N}\}. \tag{20}
$$

We pick the N store locations uniformly at random from $[0, 10]^2$. Let the unit cost t_{ij} to transport demand from location i to j be the Euclidean distance if $i \neq j$, and $t_{ii} = 0$, $i, j \in [N]$. The storage cost per unit is $c_i = 20$, $i \in [N]$, and the capacity of each store is 20 (i.e., $X = \{\mathbf{x} \in \mathbb{R}_+^{N_1} \mid \mathbf{x} \leqslant \mathbf{20}\}$). The numerical settings here are adopted from Bertsimas and de Ruiter (2016). In Table 1, we illustrate the effectiveness of our procedure introduced above. To utilize the effectiveness of RCI procedure, we repeatedly perform the following procedure: after eliminating an adjustable variable via Algorithm 1, we solve (19) for each constraint and

Table 1. Removing Redundant Constraints for Lot Sizing on a Network

#Elim.	0	5	7	9	12	13	16	19	22	25
$N=3$										
FME	13	10	21	126	—	—	—	—	—	—
Before	13	10	14	17	—	—	—	—	—	—
After	13	9	11	11	—	—	—	—	—	—
Time (s)	0	0.8	1.7	2.9	—	—	—	—	—	—
$N=4$										
FME	21	17	20	60	43,594	*	*	—	—	—
Before	21	17	18	26	75	92	102	—	—	—
After	21	17	18	23	31	33	36	—	—	—
Time (s)	0	0.6	1.9	3.6	10.2	13.5	24.3	—	—	—
$N=5$										
FME	31	26	26	37	1,096	12,521	*	*	*	*
Before	31	26	26	31	76	108	486	697	869	750
After	31	26	25	27	46	54	82	101	116	127
Time (s)	0	0	1.0	3.1	9.0	13.0	54.0	137.3	247.8	346.7
$N=10$										
FME	111	106	104	102	101	104	165	31,560	*	*
Before	111	106	104	102	101	102	125	398	1,359	*
After	111	106	104	102	100	102	116	180	343	*
Time (s)	0	0	0	0	3.6	3.7	4.7	24.3	624.3	*

Notes. We use "#Elim." to denote the number of eliminated adjustable variables, "FME" denotes the number of constraints from Algorithm 1, "Before" and "After" are the number of constraints from applying Fourier–Motzkin elimination and RCI alternately, and "Time" records the total time (in seconds) needed to detect and remove the redundant constraints thus far. All numbers reported in this table are the average of 10 replications. Here, the dash (—) stands for not applicable, and the asterisk (*) means out of memory for the current computer.

remove the constraint from the system if it is redundant. The computations reported in Table 1 were carried out with Gurobi 6.5 (Gurobi Optimization 2016) on an Intel i5-2400 3.10 GHz Windows 7 computer with 4 GB of RAM. The modeling was done using the modeling language CVX within MATLAB 2015b. Table 1 shows that the RCI procedure is very effective in removing redundant constraints for the lot-sizing problem. For instance, when $N = 4$, on average, after 12 adjustable variables are eliminated, our proposed procedure leads to merely 31 constraints, whereas only using Algorithm 1 without RCI would result in 43,594 constraints, and the total time needed for detecting and removing the redundant constraints thus far is 10.2 seconds. Note that Time is equal to 0 if the number of eliminated adjustable variables #Elim. is less than or equal to N. This is because we first eliminate the adjustable variables that have transport costs, $t_{ii} = 0$, $i \in [N]$. □

5. Extension to Adjustable Distributionally Robust Optimization

Problem (1) may seem conservative, as it does not exploit distributional knowledge of the uncertainties that may be available. It has recently been shown in Bertsimas et al. (2018) that by adopting the lifted conic representable ambiguity set of Wiesemann et al. (2014), problem (1) is also capable of modeling an ADRO problem,

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & \mathbf{c}'\mathbf{x} + \sup_{\mathbb{P}\in\mathbb{F}} \mathrm{E}_{\mathbb{P}}(\mathbf{v}'\mathbf{y}(\tilde{\mathbf{z}})) \\ \text{s.t.} \quad & \mathbf{A}(\mathbf{z})\mathbf{x} + \mathbf{B}\mathbf{y}(\mathbf{z}) \geqslant \mathbf{d}(\mathbf{z}), \quad \forall\, \mathbf{z} \in \mathcal{W}, \\ & \mathbf{x} \in X, \quad \mathbf{y} \in \mathcal{R}^{I_1, N_2}, \end{aligned} \tag{21}$$

where $\tilde{\mathbf{z}}$ is now a random variable with a conic representable support set $\mathcal{W}$, and its probability distribution is an element from the ambiguity set $\mathbb{F}$, given by

$$\mathbb{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_1}) \;\middle|\; \begin{array}{l} \mathbb{E}_{\mathbb{P}}(\mathbf{G}\tilde{\mathbf{z}}) \leqslant \boldsymbol{\mu} \\ \mathbb{P}(\tilde{\mathbf{z}} \in \mathcal{W}) = 1 \end{array} \right\},$$

with parameters $\mathbf{G} \in \mathbb{R}^{L_1 \times I_1}$ and $\boldsymbol{\mu} \in \mathbb{R}^{L_1}$. For convenience and without loss of generality, we have incorporated the auxiliary random variable defined in Bertsimas et al. (2018), Wiesemann et al. (2014) as part of $\tilde{\mathbf{z}}$, and we refer interested readers to their papers regarding the modeling capabilities of such an ambiguity set. Under Slater's condition—that is, the relative interior of $\{\mathbf{z} \in \mathcal{W} : \mathbf{G}\mathbf{z} \leqslant \boldsymbol{\mu}\}$ is nonempty—by introducing new here-and-now decision variables r and $\mathbf{s}$, Bertsimas et al. (2018) reformulate (21) into the following equivalent two-stage ARO problem:

$$\min_{(\mathbf{x},r,\mathbf{s})\in\bar{\mathcal{X}}} \mathbf{c}'\mathbf{x} + r + \mathbf{s}'\boldsymbol{\mu},$$

where

$$\bar{\mathcal{X}} = \left\{ (\mathbf{x}, r, \mathbf{s}) \in X \times \mathbb{R} \times \mathbb{R}_+^{L_1} \;\middle|\; \begin{array}{l} \exists\, \mathbf{y} \in \mathcal{R}^{I_1, N_2}: \\ r + \mathbf{s}'(\mathbf{G}\mathbf{z}) \geqslant \mathbf{v}'\mathbf{y}(\mathbf{z}), \\ \quad \forall\, \mathbf{z} \in \mathcal{W} \\ \mathbf{A}(\mathbf{z})\mathbf{x} + \mathbf{B}\mathbf{y}(\mathbf{z}) \geqslant \mathbf{d}(\mathbf{z}), \\ \quad \forall\, \mathbf{z} \in \mathcal{W} \end{array} \right\}.$$

We can now apply our approach to solve the above problem. In Section 7.2, we show that our approach significantly improves the solutions obtained in Bertsimas et al. (2018).

6. Generalization to Multistage Problems

The order of events in multistage ARO problems is as follows: The here-and-now decisions $\mathbf{x}$ are made before any uncertainty is realized, and then the uncertain parameters $\mathbf{z}_{\mathcal{S}^i}$ are revealed in the later stages, where $i \in [N_2]$ and $\mathcal{S}^i \subseteq [I_1]$. We make the decision $y_i \in \mathcal{R}^{|\mathcal{S}^i|,1}$ with the benefit of knowing $\mathbf{z}_{\mathcal{S}^i}$ but with no other knowledge of the uncertain parameters $\mathbf{z}_{\backslash\mathcal{S}^i}$ to be revealed later. We assume the *information sets* $\mathcal{S}^i \subseteq [I_1]$, $i \in [N_2]$, satisfy the following nesting condition.

Definition 2. For all $i, j \in [N_2]$, we have either $\mathcal{S}^i \subseteq \mathcal{S}^j$, $\mathcal{S}^j \subseteq \mathcal{S}^i$, or $\mathcal{S}^i \cap \mathcal{S}^j = \varnothing$.

This nesting condition is a natural assumption in multistage problems, which simply ensures our knowledge about uncertain parameters is nondecreasing over time. For example, the information sets $\mathcal{S}^1 \subseteq \mathcal{S}^2 \cdots \subseteq \mathcal{S}^{N_2} \subseteq [I_1]$ satisfy this condition. Dependencies between uncertain parameters both within and across stages can be modelled in the uncertainty set $\mathcal{W}$. The feasible set of a multistage ARO problem is as follows:

$$\begin{aligned} \mathcal{X} = \big\{ \mathbf{x} \in X \mid \exists\, y_i \in \mathcal{R}^{|\mathcal{S}^i|,1},\ \forall\, i \in [N_2]: \\ \mathbf{A}(\mathbf{z})\mathbf{x} + \mathbf{B}\mathbf{y}(\mathbf{z}) \geqslant \mathbf{d}(\mathbf{z}),\ \forall \mathbf{z} \in \mathcal{W} \big\}, \end{aligned} \tag{22}$$

where y_i is the ith element of $\mathbf{y}$, and $\mathbf{y}(\mathbf{z}) = [y_1(\mathbf{z}_{\mathcal{S}^1}), \ldots, y_{N_2}(\mathbf{z}_{\mathcal{S}^{N_2}})]' \in \mathcal{R}^{|\mathcal{S}^{N_2}|, N_2}$. While this process of decision making across stage is simple enough to state, modeling these *nonanticipativity restrictions*—that is, where a decision made now cannot be made by using exact knowledge of the later stages—is the primary complication that we address in this section as we extend our approach to the multistage case.

We propose a straightforward modification of Algorithm 1 to incorporate the nonanticipativity restrictions. Suppose the nesting condition is satisfied; we first eliminate y_l in (22) via FME, where $l = \arg\max_{i\in[N_2]} |\mathcal{S}^i|$. Similar to step 2 of Algorithm 1, we have the following constraints: there exists $y_i \in \mathcal{R}^{|\mathcal{S}^i|,1}$ for all $i \in [N_2]\backslash\{l\}$,

$$\begin{aligned} f_j(\bar{\mathbf{z}}) + \mathbf{g}_j'(\bar{\mathbf{z}})\mathbf{x} + \mathbf{h}_j'\mathbf{y}_{\backslash\{l\}}(\mathbf{z}) \geqslant f_i(\mathbf{z}) + \mathbf{g}_i'(\mathbf{z})\mathbf{x} + \mathbf{h}_i'\mathbf{y}_{\backslash\{l\}}(\mathbf{z}), \\ \forall\, (\mathbf{z}, \bar{\mathbf{z}}) \in \bar{\mathcal{W}} \text{ if } b_{il} > 0 \text{ and } b_{jl} < 0, \end{aligned} \tag{23}$$

$$0 \geqslant f_k(\mathbf{z}) + \mathbf{g}_k'(\mathbf{z})\mathbf{x} + \mathbf{h}_k'\mathbf{y}_{\setminus\{l\}}(\mathbf{z}), \quad \forall\, \mathbf{z} \in \mathcal{W} \text{ if } b_{kl} = 0, \quad (24)$$

where $\bar{\mathcal{W}} = \{(\mathbf{z}, \bar{\mathbf{z}}) \in \mathbb{R}^{2I_1} \mid \mathbf{z} \in \mathcal{W}, \bar{\mathbf{z}} \in \mathcal{W}, \mathbf{z}_{\mathcal{S}^l} = \bar{\mathbf{z}}_{\mathcal{S}^l}\}$ is an *augmented uncertainty set*. Because of the nonanticipitivity restrictions, the adjustable variable y_l only depends on $\mathbf{z}_{\mathcal{S}^l}$. The augmented uncertainty set $\bar{\mathcal{W}}$ enforces the constraints containing y_l to share the same information $\mathbf{z}_{\mathcal{S}^l}$, but the unrevealed $\mathbf{z}_{\setminus\mathcal{S}^l}$ are not necessarily the same across constraints. One simple yet crucial observation is that the nesting condition implies $\mathbf{y}_{\setminus\{l\}}(\bar{\mathbf{z}}) = \mathbf{y}_{\setminus\{l\}}(\mathbf{z})$ for all $(\mathbf{z}, \bar{\mathbf{z}}) \in \bar{\mathcal{W}}$. Hence, on the left-hand side of the inequalities (23), we have $\mathbf{y}_{\setminus\{l\}}(\mathbf{z})$ instead of $\mathbf{y}_{\setminus\{l\}}(\bar{\mathbf{z}})$. One can now update $[N_2]$ to $[N_2]\setminus\{l\}$ and further eliminate the remaining adjustable variables analogously.

7. Numerical Experiments

In this section, we evaluate the performance of our FME approach on an ARO problem and an ADRO problem. First, we further investigate the lot-sizing problem discussed in Examples 1 and 2. Then, we consider a medical appointment scheduling problem where the distributional knowledge of the uncertain consultation time of the patients is partially known.

7.1. Lot Sizing on a Network

Let us again consider P in Example 1 with the same parameter setting as in Example 2. From (11), one can write the equivalent dual formulation,

$$\begin{aligned}
\min_{\mathbf{x},\boldsymbol{\lambda},\tau}\ & \mathbf{c}'\mathbf{x} + \tau \\
\text{s.t.}\ & \omega_0\tau - 20\sqrt{N}\lambda_0(\boldsymbol{\omega}) + \sum_{i\in[N]}(\omega_i x_i - 20\lambda_i(\boldsymbol{\omega})) \geqslant 0, \quad \forall\, \boldsymbol{\omega} \in \mathcal{U}, \\
& \lambda_0(\boldsymbol{\omega}) + \lambda_i(\boldsymbol{\omega}) \geqslant \omega_i, \quad \forall\, \boldsymbol{\omega} \in \mathcal{U},\ i \in [N], \\
& \mathbf{0} \leqslant \mathbf{x} \leqslant \mathbf{20}, \\
& \boldsymbol{\lambda}(\boldsymbol{\omega}) \geqslant \mathbf{0}, \quad \boldsymbol{\lambda} \in \mathcal{R}^{N+1,N+1},
\end{aligned} \qquad \text{(D)}$$

with the dual uncertainty set,

$$\mathcal{U} = \{\boldsymbol{\omega} \in \mathbb{R}_+^{N+1} \mid -t_{ij}\omega_0 + \omega_i + \omega_j \leqslant 0,\ \mathbf{1}'\boldsymbol{\omega} = 1,\ \forall\, i, j \in [N]: i \neq j\}.$$

Note that because of the existence of $\sum_{i\in[N]}\omega_i x_i$ in the first constraint of (D), the uncertainties are not columnwise. The RCI procedure proposed in Section 4 does not detect any redundant constraint. Hence, we only

Table 2. Lot Sizing on a Network for $N \in \{5, 10\}$

#Elim.	1	11	15	19	22	25	—
$N = 5$							
P							
RCI	30	37	75	101	116	127	—
Gap%	3.3	2.9	1.7	0.7	0.1	0	—
TTime (s)	0.1	12.9	58.3	223.2	394.3	550.3	—
#Elim.	1	2	3	4	5	6	—
$N = 5$							
D							
FME	11	10	13	19	33	272	—
Gap%	3.3	2.8	2.3	1	0.2	0	—
Time (s)	0.1	0.1	0.1	0.1	0.1	0.1	—
#Elim.	1	12	17	19	21	22	100
$N = 10$							
P							
RCI	110	100	133	180	276	343	*
Gap%	6.0	6.0	6.0	5.9	5.8	5.7	*
TTime (s)	0.1	14.8	52.6	100.7	987.7	1,639.8	*
#Elim.	1	5	7	8	9	10	11
$N = 10$							
D							
FME	21	43	135	261	515	1,025	149,424
Gap%	6	4.6	2.6	1.8	0.8	0.2	*
Time (s)	0.1	0.2	0.4	0.9	1.9	4.6	*

Notes. We use "#Elim." to denote the number of eliminated adjustable variables; "RCI" is the number of resulting constraints from first applying Algorithm 1 and then the RCI procedure; "FME" denotes the number of constraints from Algorithm 1; "Gap%" denotes the average optimality gap (in percent) of 10 replications that is, for a candidate solution *sol.*, the gap is *sol.* − OPT/OPT, where OPT denotes the optimal objective value; "Time" records time (in seconds) needed to solve the corresponding optimization problem; and "TTime" reports the total time (in seconds) needed to remove the redundant constraints and solve the optimization problem.

Table 3. Lot Sizing on a Network for $N \in \{15, 20, 30\}$

#Elim.	1	2	3	4	5	6	7	8	9	10	11
$N = 15$											
FME	31	31	33	39	53	83	145	271	525	1,035	2,057
Red.%	0	−0.1	−0.4	−0.5	−0.7	−0.9	−1.6	−1.9	−2.2	−2.8	−3.4
Time (s)	0.3	0.3	0.3	0.4	0.5	1	1.5	3.6	8.9	25.2	125.1
$N = 20$											
FME	41	41	43	49	63	93	155	281	535	1,045	2,067
Red.%	0	−0.1	−0.2	−0.3	−0.4	−0.5	−0.9	−1	−1.2	−1.5	−1.8
Time (s)	0.6	0.5	0.6	0.8	1.3	2.5	3.5	10.1	36.1	67.7	206.2
$N = 30$											
FME	61	61	63	69	83	113	175	301	555	1,065	2,087
Red.%	0	0	−0.1	−0.2	−0.2	−0.3	−0.4	−0.5	−0.6	−0.8	*
Time (s)	2.2	2.2	2.6	3.4	5.8	15.0	54.6	55.7	214.5	522.2	*

Notes. Here, the asterisk (*) indicates that the average computation time exceeded the 10-minute threshold. We use "#Elim." to denote the number of eliminated adjustable variables; "FME" denotes the number of constraints from Algorithm 1; "Red.%" denotes the average cost reduction (in percent) of the approximated solution via LDRs (without constraint elimination) of 10 replications—that is, for a candidate solution *sol.*, the Red.% is *sol.* − LDR/LDR; and "Time" records the time (in seconds) needed to solve the corresponding optimization problem.

apply Algorithm 1 (without RCI) for (D). Here, the dimensions of adjustable variables in primal and dual formulations are significantly different—that is, the number of adjustable variables in the dual formulation (D) is $N+1$, whereas in the primal formulation P, it is N^2. One may expect that it is more effective to eliminate adjustable variables via Algorithm 1 in (D) than in P. We show via the following numerical experiments that it is indeed the case.

7.1.1. Numerical Study. Table 2 shows that, throughout all the experiments, solutions converge to optimality faster for (D) than for P. Hence, in Table 3, we focus on the formulation (D) for larger instances (e.g., $N \in \{15, 20, 30\}$). It shows that eliminating a subset of adjustable variables first (taking into account the computational limitation) and then solving the reformulation with LDRs leads to better solutions.

Note that the optimal objective values used in Table 2 are computed by enumerating all the vertices of the budget uncertainty set (20). For $N \leqslant 10$, the problems can be solved in five seconds on average. We also investigate the effect of the sequence in which to eliminate the adjustable variables. We observe no clear effect on the results of P if a different eliminating sequence is used. However, if we first eliminate λ_0 in (D), the number of resulting constraints increases much faster than first eliminating λ_i, $i \in [N]$. Hence, if λ_0 is eliminated first, we can only eliminate fewer adjustable variables before our computational limit is reached, which results in poorer approximations than the ones that are reported in Table 2. We suggest to first eliminate the adjustable variables that produce the smallest number of constraints (which can be easily computed before the elimination; see Section 2), such that we eliminate as many adjustable variables as possible while keeping the problem size at its minimal size.

7.2. Medical Appointment Scheduling

For the second application, we consider a medical appointment scheduling problem where patients arrive at their stipulated schedule and may have to wait in a queue to be served by a physician. The patients' consultation times are uncertain and their arrival schedules are determined at the first stage, which can influence the waiting times of the patients and the overtime of the physician. This problem is studied in Kong et al. (2013), Mak et al. (2014) and Bertsimas et al. (2018).

The problem setting here is adopted from Bertsimas et al. (2018). We consider N patients arriving in sequence with their indices $j \in [N]$ and the uncertain consultation times are denoted by $\tilde{z}_j$, $j \in [N]$. We let the first-stage decision variable, x_j, represent the inter-arrival time between patient j to the adjacent patient $j+1$ for $j \in [N-1]$, and we use x_N to denote the time between the arrival of the last patient and the scheduled completion time for the physician before overtime commences. The first patient will be scheduled to arrive at the starting time of zero, and subsequent patients i, $i \in [N], i \geqslant 2$ will be scheduled to arrive at $\sum_{j \in [i-1]} x_j$. Let T denote the scheduled completion time for the physician before overtime commences. In describing the uncertain consultation times, we consider the following partial cross-moment ambiguity set:

$$\mathbb{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^N \times \mathbb{R}^{N+1}) \;\middle|\; \begin{array}{l} \mathbb{E}_{\mathbb{P}}(\tilde{\mathbf{z}}) = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}(\tilde{u}_i) \leqslant \phi_i, \quad \forall\, i \in [N+1] \\ \mathbb{P}((\tilde{\mathbf{z}}, \tilde{\mathbf{u}}) \in \mathcal{W}) = 1 \end{array} \right\},$$

where

$$\mathcal{W} = \left\{ (\mathbf{z}, \mathbf{u}) \in \mathbb{R}^N \times \mathbb{R}^{N+1} \;\middle|\; \begin{array}{l} \mathbf{z} \geqslant \mathbf{0} \\ (z_i - \mu_i)^2 \leqslant u_i, \quad \forall\, i \in [N] \\ \left(\sum_{i \in [N]} (z_i - \mu_i) \right)^2 \leqslant u_{N+1} \end{array} \right\}.$$

Note that the introduction of the axillary random variable $\tilde{\mathbf{u}}$ in the ambiguity set is first introduced in Wiesemann et al. (2014) to obtain tractable formulations. Subsequently, Bertsimas et al. (2018) show that by incorporating it in LDRs, we could greatly improve the solutions to the adjustable distributionally robust optimization problem. A common decision criterion in the medical appointment schedule is to minimize the expected total cost of patients waiting and physician overtime, where the cost of a patient waiting is normalized to one per unit delay and the physician's overtime cost is γ per unit delay. The optimal arrival schedule $\mathbf{x}$ can be determined by solving the following two-stage adjustable distributionally robust optimization problem:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \ & \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}\left(\sum_{i\in[N]} y_i(\tilde{\mathbf{z}},\tilde{\mathbf{u}}) + \gamma y_{N+1}(\tilde{\mathbf{z}},\tilde{\mathbf{u}})\right) \\ \text{s.t.} \ & y_i(\mathbf{z},\mathbf{u}) - y_{i-1}(\mathbf{z},\mathbf{u}) + x_{i-1} \geqslant z_{i-1}, \\ & \qquad \forall (\mathbf{z},\mathbf{u})\in\mathscr{W},\ \forall i\in\{2,\ldots,N+1\}, \\ & \mathbf{y}(\mathbf{z},\mathbf{u}) \geqslant \mathbf{0}, \qquad \forall (\mathbf{z},\mathbf{u})\in\mathscr{W}, \\ & \sum_{i\in[N]} x_i \leqslant T, \\ & \mathbf{x}\in\mathbb{R}^N_+, \quad \mathbf{y}\in\mathscr{R}^{I_1+I_2,N+1}, \end{aligned} \tag{25}$$

where y_i denotes the waiting time of patient $i, i\in[N]$, and y_{N+1} represents the overtime of the physician. Since $\mathscr{W}$ is clearly not polyhedral, the reformulation technique of Bertsimas and de Ruiter (2016) cannot be applied here. As in Bertsimas et al. (2018), we use ROC to formulate the problem via LDRs, where the adjustable variables $\mathbf{y}$ are affinely in both $\mathbf{z}$ and $\mathbf{u}$, and solve it using CPLEX 12.6. ROC is developed in C++ programming language, and we refer readers to http://www.meilinzhang.com/software for more information.

7.2.1. Numerical Study. The numerical settings of our computational experiments are similar to Bertsimas et al. (2018). We have $N = 8$ jobs, and the unit overtime cost is $\gamma = 2$. For each job $i\in[N]$, we randomly select μ_i based on uniform distribution over $[30,60]$ and $\sigma_i = \mu_i\cdot\epsilon$ where ϵ is randomly selected based on uniform distribution over $[0,0.3]$. The uncertain job completion times are independently distributed, and hence we have $\phi^2 = \sum_{i=1}^N \sigma_i{}^2$. The evaluation period T depends on instance parameters as follows:

$$T = \sum_{i=1}^N \mu_i + 0.5\sqrt{\sum_{i=1}^N \sigma_i{}^2}.$$

We consider 9 reformulations of problem (25), in which 1–9 adjustable variables are eliminated, with 10 randomly generated uncertainty sets. As shown in Table 4, the RCI procedure effectively removes the redundant constraints in the reformulations. After 92.3 seconds of preprocessing, all 9 adjustable variables are eliminated, which ends up with only 255 constraints, whereas only using Algorithm 1 without RCI leads to so many constraints that our computer is out of memory. Although computing the reformulations can be time consuming, we only need to compute the reformulations once, because our reformulation procedure via Algorithm 1 and RCI is independent from the uncertainty set of problem (25). For the 10 randomly generated uncertainty sets, the average optimality gap of the solutions obtained in Bertsimas et al. (2018) is 12.8%. Our approach reduces the optimality gap to zero when more adjustable variables are eliminated. Since the size of this problem is relatively small, the computational times for all the instances in Table 4 are less than two seconds. Finally, similar to the primal formulation of the lot-sizing problem, we observe no clear effect on the obtained results if different eliminating sequences are

Table 4. Appointment Scheduling for $N = 8$

# Elim.	0	1	2	3	4	5	6	7	8	9
FME	18	17	17	20	37	132	731	5,050	40,329	*
Before	18	17	17	20	29	52	107	234	521	1,152
After	18	17	17	18	21	28	43	74	137	255
Time (s)	0	0.7	1.0	1.1	1.5	2.8	5.8	12.9	30.7	74.9
Obj.	155	155	155	155	155	152	148	145	142	138
Gap%	12.8	12.8	12.8	12.7	12.5	10.5	7.6	5.6	3.3	0
Min. Gap%	10.4	10.4	10.4	10.4	10.3	9.7	6.4	4.8	2.5	0
Max. Gap%	14.6	14.6	14.6	14.5	13.5	11.5	8.1	6.2	3.6	0

Notes. Here, the asterisk (*) means out of memory for the current computer. We use "#Elim." to denote the number of eliminated adjustable variables; "FME" denotes the number of constraints from Algorithm 1; "Before" and "After" are the number of constraints from applying Algorithm 1 and RCI alternately; "Time" records the total time (in seconds) needed to detect and remove the redundant constraints thus far; "Obj." denotes the average objective value obtained from solving (25) via LDRs; and "Min. Gap%," "Max. Gap%," and "Gap%" record the minimum, maximum, and average optimality gap (in %) of 10 replications, respectively—that is, for a candidate solution *sol.*, the gap is *sol.* − OPT/OPT, where OPT denotes the optimal objective value. All numbers reported in the last four rows are the average of 10 replications.

considered. Furthermore, the number of constraints after the eliminations and the RCI procedures remains unchanged for problem (25) if different eliminating sequences are used.

8. Conclusions

We propose a generic FME approach for solving ARO problems with fixed recourse to optimality. Through the lens of FME, we characterize the structures of the ODRs for a broad class of ARO problems. We extend the approach of Bertsimas et al. (2018) for ADRO problems. Via numerical experiments, we show that for small-sized ARO problems, our approach finds the optimal solution, and for moderate- to large-sized instances, we successively improve the approximated solutions obtained from LDRs.

On a theoretical level, one immediate future research direction would be to characterize the structures of the ODRs for multistage problems; for example, see Bertsimas et al. (2010), Iancu et al. (2013). Another potential direction would be to extend our FME approach to ARO problems with integer adjustable variables or nonfixed recourse.

On a numerical level, we would like to investigate the performance of Algorithm 1 with finite adaptability approaches or other decision rules on solving ARO problems. Moreover, many researchers have proposed alternative approaches for identifying redundant constraints in LP problems. For instance, Huynh et al. (1992) discuss the efficiency of three alternative procedures for computing polytopic projections and introduce a new RCI method; Paulraj and Sumathi (2010) compare the efficiency of five RCI methods. Another potential direction would be to adapt and combine the existing alternative procedures to further improve the efficiency of our proposed approach.

Acknowledgments

The authors are grateful to the associate editor and two anonymous referees for valuable comments on an earlier version of the paper. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Singapore Ministry of Education or the Singapore Government.

Appendix. Proof of the Dual Formulation (10) for Problem (1)

We first represent problem (1) in the equivalent form:

$$\min_{\mathbf{x}\in X}\max_{\mathbf{z}\in\mathscr{W}}\min_{\mathbf{y}}\{\mathbf{c}'\mathbf{x}\mid \mathbf{A}(\mathbf{z})\mathbf{x}+\mathbf{B}\mathbf{y}\geqslant \mathbf{d}(\mathbf{z})\}.$$

Because of strong duality, we obtain the following reformulation by dualizing $\mathbf{y}$:

$$\min_{\mathbf{x}\in X}\max_{\mathbf{z}\in\mathscr{W},\,\boldsymbol{\omega}\in\mathbb{R}^M_+}\{\mathbf{c}'\mathbf{x}\mid \boldsymbol{\omega}'(\mathbf{d}(\mathbf{z})-\mathbf{A}(\mathbf{z})\mathbf{x})\leqslant 0,\mathbf{B}'\boldsymbol{\omega}=\mathbf{0}\}.$$

Similarly, by further dualizing $\mathbf{z}\in\mathscr{W}_{\text{poly}}=\{\mathbf{z}\in\mathbb{R}^{I_1}\mid \exists\mathbf{v}\in\mathbb{R}^{I_2}: \mathbf{P}'\mathbf{z}+\mathbf{Q}'\mathbf{v}\leqslant\boldsymbol{\rho}\}$, we have

$$\min_{\mathbf{x}\in X}\max_{\boldsymbol{\omega}\in\mathbb{R}^M_+}\min_{\boldsymbol{\lambda}\geqslant 0}\left\{\mathbf{c}'\mathbf{x}\;\middle|\;\begin{array}{l}\boldsymbol{\omega}'(\mathbf{d}^0-\mathbf{A}^0\mathbf{x})+\boldsymbol{\rho}'\boldsymbol{\lambda}\leqslant\mathbf{0}\\ \mathbf{p}_i'\boldsymbol{\lambda}=(\mathbf{d}^i-\mathbf{A}^i\mathbf{x})'\boldsymbol{\omega},\quad\forall i\in[I_1]\\ \mathbf{Q}\boldsymbol{\lambda}=\mathbf{0},\quad \mathbf{B}'\boldsymbol{\omega}=\mathbf{0}\end{array}\right\},$$

where $\mathbf{p}_i\in\mathbb{R}^{I_1}$, $i\in[I_1]$, is the ith row vector of matrix $\mathbf{P}$, which can be represented equivalently as

$$\min_{\mathbf{x}\in X}\left\{\mathbf{c}'\mathbf{x}\;\middle|\;\exists\boldsymbol{\lambda}\in\mathscr{R}^{M,K}:\begin{array}{r}\boldsymbol{\omega}'(\mathbf{A}^0\mathbf{x}-\mathbf{d}^0)-\boldsymbol{\rho}'\boldsymbol{\lambda}(\boldsymbol{\omega})\geqslant\mathbf{0},\\ \forall\boldsymbol{\omega}\in\mathscr{U}\\ \mathbf{p}_i'\boldsymbol{\lambda}(\boldsymbol{\omega})=(\mathbf{d}^i-\mathbf{A}^i\mathbf{x})'\boldsymbol{\omega},\\ \forall\boldsymbol{\omega}\in\mathscr{U},\ \forall i\in[I_1]\\ \mathbf{Q}\boldsymbol{\lambda}(\boldsymbol{\omega})=\mathbf{0},\quad\boldsymbol{\lambda}(\boldsymbol{\omega})\geqslant\mathbf{0},\\ \forall\boldsymbol{\omega}\in\mathscr{U}\end{array}\right\},$$

where $\mathscr{U}=\{\boldsymbol{\omega}\in\mathbb{R}^M_+\mid\mathbf{B}'\boldsymbol{\omega}=\mathbf{0}\}$. □

References

Ardestani-Jaafari A, Delage E (2016a) Linearized robust counterparts of two-stage robust optimization problems with applications in operations management. Working paper, HEC Montreal, Montreal. http://optimization-online.org/DB_FILE/2016/03/5388.pdf.

Ardestani-Jaafari A, Delage E (2016b) Robust optimization of sums of piecewise linear functions with application to inventory problems. *Oper. Res.* 64(2):474–494.

Bemporad A, Borrelli F, Morari M (2003) Min-max control of constrained uncertain discrete-time linear systems. *IEEE Trans. Automatic Control* 48(9):1600–1606.

Ben-Ameur W, Wang G, Ouorou A, Zotkiewicz M (2017) Multipolar robust optimization. *EURO J. Comput. Optim.*, https://doi.org/10.1007/s13675-017-0092-4.

Ben-Tal A, Nemirovski A (1998) Robust convex optimization. *Math. Oper. Res.* 23(4):769–805.

Ben-Tal A, Nemirovski A (1999) Robust solutions of uncertain linear programs. *Oper. Res. Lett.* 25(1):1–13.

Ben-Tal A, Nemirovski A (2000) Robust solutions of linear programming problems contaminated with uncertain data. *Math. Programming* 88(3):411–424.

Ben-Tal A, den Hertog D, Vial JP (2015) Deriving robust counterparts of nonlinear uncertain inequalities. *Math. Programming* 149(1):265–299.

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) *Robust Optimization*, Princeton Series in Applied Mathematics (Princeton University Press, Princeton, NJ).

Ben-Tal A, El Housni O, Goyal V (2016) A tractable approach for designing piecewise affine policies in dynamic robust optimization. Working paper, Technion–Israel Institute of Technology, Haifa. http://optimization-online.org/DB_FILE/2016/07/5557.pdf.

Ben-Tal A, Goryashko A, Guslitzer E, Nemirovski A (2004) Adjustable robust solutions of uncertain linear programs. *Math. Programming* 99(2):351–376.

Bertsimas D, Brown D (2009) Constructing uncertainty sets for robust linear optimization. *Oper. Res.* 57(6):1483–1495.

Bertsimas D, Caramanis C (2010) Finite adaptability for linear optimization. *IEEE Trans. Automatic Control* 55(12):1751–2766.

Bertsimas D, de Ruiter F (2016) Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds. *INFORMS J. Comput.* 28(3):500–511.

Bertsimas D, Dunning I (2016) Multistage robust mixed integer optimization with adaptive partitions. *Oper. Res.* 64(4):980–998.

Bertsimas D, Georghiou A (2015) Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Oper. Res.* 63(3):610–627.

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

Bertsimas D, Sim M (2004) The price of robustness. *Oper. Res.* 52(1):35–53.

Bertsimas D, Tsitsiklis J (1997) *Introduction to Linear Optimization* (Athena Scientific, Belmont, MA).

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

Bertsimas D, Iancu D, Parrilo P (2010) Optimality of affine policies in multistage robust optimization. *Math. Oper. Res.* 35(2):363–394.

Bertsimas D, Iancu D, Parrilo P (2011) A hierarchy of near-optimal policies for multistage adaptive optimization. *IEEE Trans. Automatic Control* 56(12):2809–2824.

Bertsimas D, Sim M, Zhang M (2018) A practically efficient approach for solving adaptive distributionally robust linear optimization problems. *Management Sci.*. Forthcoming.

Birge JR, Louveaux F (1997) *Introduction to Stochastic Programming* (Springer, New York).

Breton M, El Hachem S (1995) Algorithms for the solution of stochastic dynamic minimax problems. *Comput. Optim. Appl.* 4(4): 317–345.

Caron R, McDonald J, Ponic C (1989) A degenerate extreme point strategy for the classification of linear constraints as redundant or necessary. *J. Optim. Theory* 62(2):225–237.

Chen W, Sim M (2009) Goal-driven optimization. *Oper. Res.* 57(2): 342–357.

Chen X, Zhang Y (2009) Uncertain linear programs: Extended affinely adjustable robust counterparts. *Oper. Res.* 57(6):1469–1482.

Chen X, Sim M, Sun P (2007) A robust optimization perspective on stochastic programming. *Oper. Res.* 55(6):1058–1071.

Chen X, Sim M, Sun P, Zhang J (2008) A linear decision-based approximation approach to stochastic programming. *Oper. Res.* 56(2):344–357.

Dantzig G (1963) *Linear Programming and Extensions* (Princeton University Press, Princeton, NJ).

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58(3):596–612.

El Ghaoui L, Lebret H (1997) Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Anal. Appl.* 18(4): 1035–1064.

El Ghaoui L, Oustry F, Lebret H (1998) Robust solutions to uncertain semidefinite programs. *SIAM J. Optim.* 9(1):33–53.

Fourier J (1826) Reported in: Analyse des travaux de l'Académie Royale des Sciences, pendant l'année 1824, Partie mathématique. *Histoire de l'Academie Royale des Sciences de l'Institut de France* 7:47–55.

Garstka S, Wets R (1974) On decision rules in stochastic programming. *Math. Programming* 7(1):117–143.

Goh J, Sim M (2010) Distributionally robust optimization and its tractable approximations. *Oper. Res.* 58(4):902–917.

Gorissen B, den Hertog D (2013) Robust counterparts of inequalities containing sums of maxima of linear functions. *Eur. J. Oper. Res.* 227(1):30–43.

Gorissen B, Ben-Tal A, Blanc H, den Hertog D (2014) Deriving robust and globalized robust solutions of uncertain linear programs with general convex uncertainty sets. *Oper. Res.* 62(3):672–679.

Gurobi Optimization (2016) Gurobi optimizer reference manual, http://www.gurobi.com.

Hanasusanto G, Kuhn D, Wiesemann W (2014) *K-adaptability in two-stage robust binary programming*. *Oper. Res.* 63(4):877–891.

Huynh T, Lassez C, Lassez J-L (1992) Practical issues on the projection of polyhedral sets. *Ann. Math. Artificial Intelligence* 6(4): 295–316.

Iancu D, Sharma M, Sviridenko M (2013) Supermodularity and affine policies in dynamic robust optimization. *Oper. Res.* 61(4): 941–956.

Kali P, Wallace S (1994) *Stochastic Programming* (John Wiley & Sons, Chichester, UK).

Kong Q, Lee C, Teo C, Zheng Z (2013) Scheduling arrivals to a stochastic service delivery system using copositive cones. *Oper. Res.* 61(3):711–726.

Kuhn D, Wiesemann W, Georghiou A (2011) Primal and dual linear decision rules in stochastic and robust optimization. *Math. Programming* 130(1):177–209.

Mak H, Rong Y, Zhang J (2014) Appointment scheduling with limited distributional information. *Management Sci.* 61(2):316–334.

Minoux M (2011) On 2-stage robust LP with RHS uncertainty: Complexity results and applications. *J. Global Optim.* 49(3):521–537.

Motzkin T (1936) *Beiträge zur Theorie der linearen Ungleichungen* (Azriel, Jerusalem, Israel).

Mutapcic A, Boyd S (2009) Cutting-set methods for robust convex optimization with pessimizing oracles. *Optim. Methods Software* 24(3):381–406.

Paulraj S, Sumathi P (2010) A comparative study of redundant constraints identification methods in linear programming problems. *Math. Problems Engrg.* 2010:Article 723402.

Popescu I (2007) Robust mean-covariance solutions for stochastic optimization. *Oper. Res.* 55(4):98–112.

Postek K, den Hertog D (2016) Multi-stage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. *INFORMS J. Comput.* 28(3):553–574.

Scarf H (1958) A min-max solution of an inventory problem. Arrow K, ed. *Studies in the Mathematical Theory of Inventory and Production* (Stanford University Press, Stanford, CA), 201–209.

See C-T, Sim M (2009) Robust approximation of multiperiod inventory management. *Oper. Res.* 58(3):583–594.

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

Shapiro A, Kleywegt A (2002) Minimax analysis of stochastic programs. *Optim. Methods Software* 17(3):523–542.

Vayanos P, Kuhn D, Rustem B (2011) Decision rules for information discovery in multi-stage stochastic programming. *Proc. 50th IEEE Conf. Decision Control Eur. Control Conf.* (IEEE, Piscataway, NJ), 7368–7373.

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

Xu G, Burer S (2018) A copositive approach for two-stage adjustable robust optimization with uncertain right-hand sides. *Comput. Optim. Appl.* 70(1):33–59.

Xu H, Mannor S (2012) Distributionally robust Markov decision processes. *Math. Oper. Res.* 37(2):288–300.

Žáčková J (1966) On minimax solution of stochastic linear programming problems. *Časopis pro Pěstování Matematiky* 91(4):423–430.

Zhen J, den Hertog D (2017) Centered solutions for uncertain linear equations. *Comput. Management Sci.* 14(4):585–610.

Zhen J, den Hertog D (2018) Computing the maximum volume inscribed ellipsoid of a polytopic projection. *INFORMS J. Comput.* 30(1):31–42.

Jianzhe Zhen is a postdoc at École Polytechnique Fédérale de Lausanne, Switzerland. His research is focused on robust optimization. He won the Student Best Paper Prize for this paper at the Computational Management Science 2017 conference in Bergamo.

Dick den Hertog is professor of business analytics and operations research at Tilburg University. His research interests cover various fields in prescriptive analytics, in particular linear and nonlinear optimization. In recent years his main focus has been on robust optimization and simulation-based optimization. In 2000 he received the EURO Best Applied Paper Award, together with Peter Stehouwer. In 2013 he was a member of the team that received the INFORMS Franz Edelman Award.

Melvyn Sim is a professor at the Department of Analytics and Operations, NUS Business School. His research interests fall broadly under the categories of decision making and optimization under uncertainty with applications in finance, supply chain management, healthcare, and engineered systems.