manuscript

Adjustable Robust Optimization via Fourier-Motzkin Elimination

Jianzhe Zhen*, Dick den Hertog
Department of Econometrics and Operations Research, Tilburg University,
j.zhen@tilburguniversity.edu, d.denhertog@tilburguniversity.edu

Melvyn Sim
NUS Business School, National University of Singapore, melvynsim@nus.edu.sg

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 enables us to solve adjustable optimization problems to optimality. We show via numerical experiments that, for small-size ARO problems our novel approach finds the optimal solution. For moderate or large-size instances, we eliminate a subset of the adjustable variables, which improves the solutions obtained from linear decision rules.

Key words: 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.

* Corresponding author.

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 due to suboptimality (see Garstka and Wets 1974). Interestingly, there is also a revival of using LDRs for solving multi-stage stochastic optimization problems (Kuhn et al. (2011)). Other nonlinear decision rules in the recent literature include, e.g., quadratic decision rules in Ben-Tal et al. (2009), 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. (2012), 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, whilst 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 (1995). 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. (2017) 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 introduce in Fourier (1826), and was rediscovered in Motzkin (1936). Via FME, we reformulate the ARO problems into their equivalent counterparts with a reduced number of adjustable variables at the expense of an increasing number of constraints. Theoretically, every ARO problem admits an equivalent static reformulation, however, one major obstacle in practice is that FME often leads to too many redundant constraints. In order to keep the resulting equivalent counterpart at its minimal size, after eliminating an adjustable variable via FME, we execute an 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 (2017) 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; 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. (2017) for ADRO problems. Via numerical experiments, we show that our approach improves the obtained solutions in Bertsimas et al. (2017).

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 of 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. (2017) for ADRO problems.

5. Via 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.

This paper is organized as follows. In §2, we introduce FME for two-stage ARO problems. §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 §4, we propose an LP-based RCI procedure to remove the redundant constraints. §5 uses our FME approach to extend the approach of Bertsimas et al. (2017) for ADRO problems. We generalize our approach to the multistage case in §6. §7 evaluates our approach numerically via lot-sizing on a network and appointment scheduling problems. §8 presents conclusions and future research.

Notations. We use $[N]$, $N \in \mathbb{N}$ to denote the set of running indices, $\{1, \dots, N\}$. We generally use bold faced characters such as $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ to represent vectors and matrices, respectively, and $\boldsymbol{x}_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|}$ to denote a vector that contains a subset $\mathcal{S} \subseteq [N]$ of components in $\boldsymbol{x}$, e.g., $x_i \in \mathbb{R}$ denotes the i-th element of $\boldsymbol{x}$. We use $(x)^+$ and $|x|$ to denote $\max\{x, 0\}$ and the absolute value of $x \in \mathbb{R}$, and $|\mathcal{S}|$ to denote the cardinality of a finite set $\mathcal{S} \subseteq [N]$. Special vectors include $\mathbf{0}$, $\mathbf{1}$ and $\boldsymbol{e}_i$ 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 tilde to denote a random variable without associating it with a particular probability distribution. We use $\tilde{\boldsymbol{z}} \in \mathbb{R}^I$ 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 over the probability distribution $\mathbb{P}$. For a support set $\mathcal{W} \subseteq \mathbb{R}^I$, $\mathbb{P}(\tilde{\boldsymbol{z}} \in \mathcal{W})$ represents the probability of $\tilde{\boldsymbol{z}}$ 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 $\boldsymbol{x} \in \mathbb{R}^{N_1}$ are decided before the realization of the uncertain parameters $\boldsymbol{z}$, and the second stage or *wait-and-see* decisions $\boldsymbol{y}$ are determined after the value of $\boldsymbol{z}$ is revealed, and $\boldsymbol{z}$ resides in a set $\mathcal{W} \subset \mathbb{R}^{I_1}$. Let us call $\boldsymbol{y}$ *adjustable variables.* With this setting, a two-stage ARO problem can be written as follows:

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

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

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

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

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

with $\boldsymbol{A}^0, \boldsymbol{A}^1, ..., \boldsymbol{A}^{I_1} \in \mathbb{R}^{M \times N_1}$ and $\boldsymbol{d}^0, \boldsymbol{d}^1, ..., \boldsymbol{d}^{I_1} \in \mathbb{R}^M$. The matrix $\boldsymbol{B} \in \mathbb{R}^{M \times N_2}$, 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. (2017) show that it is capable of modeling adjustable distributionally robust optimization (ADRO) problems. In §5, we show how to apply our approach to solve ADRO problems. We then generalize our approach to the multistage case in §6. Problem (1) is generally intractable, even if there are only right hand side uncertainties (see Minoux (2011)), because the adjustable variables $\boldsymbol{y}$ 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 $\boldsymbol{y}$ via Fourier-Motzkin elimination (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 $\boldsymbol{x} \in \mathcal{X}$. This algorithm is adapted from (Bertsimas and Tsitsiklis 1997, page 72) for polyhedral projections.

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 $\boldsymbol{y} \in \mathcal{R}^{I_1, N_2}$,*

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

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 $\boldsymbol{y} \in \mathcal{R}^{I_1, N_2}$,

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

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

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

Here, each $\boldsymbol{h}_i, \boldsymbol{h}_j, \boldsymbol{h}_k$ is a vector in $\mathbb{R}^{N_2-1}$, for a given $\boldsymbol{z}$, each f_i, f_j, f_k is a scalar, and each $\boldsymbol{g}_i, \boldsymbol{g}_j, \boldsymbol{g}_k$ 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 $\boldsymbol{y}_{\setminus\{l\}} \in \mathcal{R}^{I_1, N_2-1}$,*

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

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

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. $\mathcal{X} = \mathcal{X}_{\setminus\{l\}}$.

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

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

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

Let

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

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

From Theorem 1, one can repeatedly apply Algorithm 1 to eliminate all the linear adjustable variables $\boldsymbol{y}$ 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:

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

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 §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 $\boldsymbol{z}$ 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:

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

The transportation cost $t_{ij} = 0$, if $i = j$; $t_{ij} \ge 0$, otherwise. For $N = 2$, there are 4 adjustable variables, i.e., y_{11}, y_{12}, y_{21} and y_{22}. We apply Algorithm 1 iteratively, which leads to the following equivalent reformulation:

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

Note that we omit $\tau \geq 0$, 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 §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 $\mathcal{X}_{\setminus[N_2]}$. We can first (iteratively) eliminate a subset $\mathcal{S}\subseteq[N_2]$ of the adjustable variables in $\mathcal{X}$ till the size of the resulting description $\mathcal{X}_{\setminus\mathcal{S}}$ reaches the prescribed computational limit, and then impose some simple functions (i.e., decision rules) $\mathcal{F}^{I_1,1}\subset\mathcal{R}^{I_1,1}$ on the remaining y_i, for all $i\in[N_2]\setminus\mathcal{S}$. The feasible set becomes:

$$\widehat{\mathcal{X}}_{\setminus\mathcal{S}} = \left\{\boldsymbol{x}\in X \mid \exists \boldsymbol{y}_{\setminus\mathcal{S}}\in\mathcal{F}^{I_1,N_2-|\mathcal{S}|} : \boldsymbol{G}(\boldsymbol{z})\boldsymbol{x}+\boldsymbol{H}\boldsymbol{y}(\boldsymbol{z})\geq \boldsymbol{f}(\boldsymbol{z}) \quad \forall \boldsymbol{z}\in\mathcal{W}\right\},$$

where $\boldsymbol{G}(\boldsymbol{z})$ and $\boldsymbol{H}$ are the resulting coefficient matrices of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, and $\boldsymbol{f}(\boldsymbol{z})$ is the corresponding right-hand side vector after elimination. Since $\boldsymbol{y}\in\mathcal{F}^{I_1,N_2}\subset\mathcal{R}^{I_1,N_2}$, it follows that $\widehat{\mathcal{X}}_{\setminus\mathcal{S}}$ is a conservative (inner) approximation of $\mathcal{X}_{\setminus\mathcal{S}}$, i.e., $\widehat{\mathcal{X}}_{\setminus\mathcal{S}}\subseteq\mathcal{X}_{\setminus\mathcal{S}}$. We simply use $\widehat{\mathcal{X}}$ to denote $\widehat{\mathcal{X}}_{\setminus\emptyset}$. 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., $\widehat{\mathcal{X}}_{\setminus[N_2]}=\mathcal{X}_{\setminus[N_2]}=\mathcal{X}$.

THEOREM 2. *$\widehat{\mathcal{X}}\subseteq\widehat{\mathcal{X}}_{\setminus\mathcal{S}_1}\subseteq\widehat{\mathcal{X}}_{\setminus\mathcal{S}_2}\subseteq\mathcal{X}$, for all $\mathcal{S}_1\subseteq\mathcal{S}_2\subseteq[N_2]$.*

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

$$\mathcal{X}_{\setminus\mathcal{S}} = \left\{\boldsymbol{x}\in X \mid \exists \boldsymbol{y}_{\setminus\mathcal{S}}\in\mathcal{R}^{I_1,N_2-|\mathcal{S}|} : \boldsymbol{G}(\boldsymbol{z})\boldsymbol{x}+\boldsymbol{H}\boldsymbol{y}_{\setminus\mathcal{S}}(\boldsymbol{z})\geq \boldsymbol{f}(\boldsymbol{z}) \quad \forall \boldsymbol{z}\in\mathcal{W}\right\},$$

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

$$\begin{aligned}
\widehat{\mathcal{X}}_{\setminus\mathcal{S}} &= \left\{\boldsymbol{x}\in X \mid \exists \boldsymbol{y}_{\setminus\mathcal{S}}\in\mathcal{F}^{I_1,N_2-|\mathcal{S}|} : \boldsymbol{G}(\boldsymbol{z})\boldsymbol{x}+\boldsymbol{H}\boldsymbol{y}_{\setminus\mathcal{S}}(\boldsymbol{z})\geq \boldsymbol{f}(\boldsymbol{z}) \quad \forall \boldsymbol{z}\in\mathcal{W}\right\} \\
&= \left\{\boldsymbol{x}\in X \mid \exists \boldsymbol{y}_{\setminus\mathcal{S}}\in\mathcal{F}^{I_1,N_2-|\mathcal{S}|},\ \boldsymbol{y}_{\mathcal{S}}\in\mathcal{R}^{I_1,|\mathcal{S}|} : \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}(\boldsymbol{z})\geq \boldsymbol{d}(\boldsymbol{z}) \quad \forall \boldsymbol{z}\in\mathcal{W}\right\},
\end{aligned}$$

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

Theorem 2 shows that Algorithm 1 can be used to improve the solutions of all existing methods, which includes 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 appear linearly in the constraints. E.g., one can use Algorithm 1 to eliminate y_l in

$$\mathcal{X}^{ge} = \left\{ \boldsymbol{x} \in X \mid \exists \boldsymbol{y} \in \mathcal{R}^{I_1,N_2} : \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}_{\backslash\{l\}}, \boldsymbol{z}) + \boldsymbol{b} y_l \geq \boldsymbol{0} \quad \forall \boldsymbol{z} \in \mathcal{W} \right\},$$

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

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 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, i.e., linear decision rule (LDR), $\boldsymbol{y} \in \mathcal{F}^{I_1,N_2} = \mathcal{L}^{I_1,N_2}$, where

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

and $\boldsymbol{y}^i \in \mathbb{R}^{N_2}$, $i \in [I_1] \cup \{0\}$, are decision variables. We show that for small-size ARO problems our approach gives a static tractable counterpart of the ARO problems and finds the optimal solution. For moderate or large-size 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 $\boldsymbol{y}$ are general piecewise affine functions.*

Proof. Let us denote $\boldsymbol{x}^*$ 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 $\mathcal{S}_l = [N_2] \setminus \{l\}$. From Theorem 1, we know $\mathcal{X} = \mathcal{X}_{\setminus \mathcal{S}_l}$. The adjustable variable y_l is upper (lower) bounded by a finite number of minimum (maximum) of affine functions in $\boldsymbol{z}$, i.e.,

$$\check{f}_l(\boldsymbol{z}) \leq y_l(\boldsymbol{z}) \leq \hat{f}_l(\boldsymbol{z}) \qquad \forall \boldsymbol{z} \in \mathcal{W}, \tag{9}$$

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

$$\check{f}_l(\boldsymbol{z}) \leq \hat{f}_l(\boldsymbol{z}) \qquad \forall \boldsymbol{z} \in \mathcal{W}$$

must hold and hence $y_l(\boldsymbol{z}) = \check{f}_l(\boldsymbol{z})$ and $y_l(\boldsymbol{z}) = \hat{f}_l(\boldsymbol{z})$ 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 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, $i \in [N_2] \setminus \{l\}$, 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

$$\mathcal{W}_{poly} = \left\{ \boldsymbol{z} \in \mathbb{R}^{I_1} \ \mid \ \exists \boldsymbol{v} \in \mathbb{R}^{I_2} : \boldsymbol{P}'\boldsymbol{z} + \boldsymbol{Q}'\boldsymbol{v} \leq \boldsymbol{\rho} \right\},$$

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

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

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

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

with the *dual uncertainty set*:

$$\mathcal{U} = \left\{ \boldsymbol{\omega} \in \mathbb{R}^M_+ \ \mid \ \boldsymbol{B}'\boldsymbol{\omega} = \boldsymbol{0} \right\},$$

where $\boldsymbol{p}_i \in \mathbb{R}^{I_1}$ are the i-th row vectors of matrix $\boldsymbol{P}$ for $i \in [I_1]$. There exist auxiliary variables $\boldsymbol{v}$ in $\mathcal{W}_{poly}$. For the decision rules of Problem (1), the adjustable variables $\boldsymbol{y}$ should depend on both $\boldsymbol{z}$ and $\boldsymbol{v}$. 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 (2017) 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 $\mathcal{W}_{poly}$, i.e., a standard simplex and a box.

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

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

$$\mathcal{W}_{simplex} = \left\{ \boldsymbol{z} \in \mathbb{R}^{I_1}_+ \ \middle| \ \boldsymbol{1}'\boldsymbol{z} \leq 1 \right\}.$$

From (11), we have the following reformulation:

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

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

$$\boldsymbol{\omega}'(\boldsymbol{A}^0\boldsymbol{x}-\boldsymbol{d}^0) \geq \lambda(\boldsymbol{\omega}) \geq \left(\max_{i\in[I_1]}\{(\boldsymbol{d}^i-\boldsymbol{A}^i\boldsymbol{x})'\boldsymbol{\omega}\}\right)^+ \qquad \forall \boldsymbol{\omega}\in\mathcal{U}.$$

Hence, there exists an ODR in the form of $\lambda(\boldsymbol{\omega})=\boldsymbol{\omega}'(\boldsymbol{A}^0\boldsymbol{x}-\boldsymbol{d}^0)$, 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 $\boldsymbol{y}$ in the primal formulation (1). □

Theorem 4 coincides with the recent finding in (Ben-Ameur et al. 2016, 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 (2017) 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 $\boldsymbol{y}$ in Problem (1) with general convex uncertainty sets.

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

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

$$\mathcal{W}_{box}=\left\{\boldsymbol{z}\in\mathbb{R}^{I_1} \;\middle|\; -\boldsymbol{\rho}\leq\boldsymbol{z}\leq\boldsymbol{\rho}\right\},$$

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

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

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

$$\begin{aligned} \frac{1}{2\rho_l}\left[\boldsymbol{\omega}'(\boldsymbol{A}^0\boldsymbol{x}-\boldsymbol{d}^0)+\sum_{i\in[I_1]}\rho_i(\boldsymbol{A}^i\boldsymbol{x}-\boldsymbol{d}^i)'\boldsymbol{\omega}-\sum_{i\in[I_1]\backslash\{l\}}2\rho_i\left((\boldsymbol{A}^i\boldsymbol{x}-\boldsymbol{d}^i)'\boldsymbol{\omega}\right)^+\right] &\geq \lambda_l(\boldsymbol{\omega}) \qquad \forall\boldsymbol{\omega}\in\mathcal{U} \\ \left((\boldsymbol{A}^i\boldsymbol{x}-\boldsymbol{d}^i)'\boldsymbol{\omega}\right)^+ &\leq \lambda_l(\boldsymbol{\omega}) \qquad \forall\boldsymbol{\omega}\in\mathcal{U}. \end{aligned}$$

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

$$\mathcal{X}^D_{\backslash[I_1]} = \left\{ \boldsymbol{x} \in X \;\middle|\; \forall \boldsymbol{\omega} \in \mathcal{U} : \boldsymbol{\omega}'(\boldsymbol{A}^0\boldsymbol{x} - \boldsymbol{d}^0) \geq \sum_{i\in[I_1]} \rho_i |(\boldsymbol{d}^i - \boldsymbol{A}^i\boldsymbol{x})'\boldsymbol{\omega}| \right\}.$$

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 $\boldsymbol{\omega} \in \mathcal{U}$ 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 $\boldsymbol{y}$ 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 §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. Firstly, we give a formal definition of redundant constraints for ARO problems.

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

$$\boldsymbol{a}_i'(\boldsymbol{\zeta})\boldsymbol{x} + \boldsymbol{b}_i'\boldsymbol{y}(\boldsymbol{\zeta}) \geq d_i(\boldsymbol{\zeta}) \qquad \forall \boldsymbol{\zeta} \in \mathcal{W}, \quad \forall i \in [M]\backslash\{l\}, \tag{13}$$

then

$$\boldsymbol{a}_l'(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{b}_l'\boldsymbol{y}(\boldsymbol{z}) \geq d_l(\boldsymbol{z}) \qquad \forall \boldsymbol{z} \in \mathcal{W}, \tag{14}$$

where $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are the i-th row vectors of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively, and d_i is the i-th component of $\boldsymbol{d}$ 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 $\boldsymbol{x}$. The redundant constraint identification (RCI) procedure in Theorem 6 is inspired by Caron et al. (1989).

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

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

has nonnegative optimal objective, i.e., $Z_l^* \geq 0$.

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

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

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

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

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*

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

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

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

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

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

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

Hence, whenever $Z_l^\dagger \geq 0$, we have $Z_l^* \geq 0$, implying the l-th 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 column-wise. Specifically let $\{\boldsymbol{z}^0, ..., \boldsymbol{z}^{N_1}\}$, $\boldsymbol{z}^j \in \mathbb{R}^{I_1^j}$, $j \in [N_1] \cup \{0\}$ be a partition of the vector $\boldsymbol{z} \in \mathbb{R}^{I_1}$ into $N_1 + 1$ vectors (including empty ones) such that

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

Note that if $\boldsymbol{z}^j$, $j \in [N_1]$ 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 $x_j \geq 0$ 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 $i \in \mathcal{M}_1$, $\boldsymbol{b}_i \neq \boldsymbol{0}$ 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 $\boldsymbol{z}^j$, for all $j \in [N_2] \cup \{0\}$, specifically,

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

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

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

The l-th inequality is redundant if the optimal objective value is nonnegative. Due to the presence of products of variables (e.g., $\boldsymbol{z}^j x_j$), Problem (18) is nonconvex in $\boldsymbol{x}$ and $\boldsymbol{z}$. An equivalent convex representation of (18) can be obtained by substituting $\boldsymbol{w}^j = \boldsymbol{z}^j x_j$, $j\in\mathcal{S}$,

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

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

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

Hence, (19) is a convex optimization problem. This transformation technique is first proposed in Dantzig (1963) to solve Generalized LPs. 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 column-wise 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 pre-computed offline and used to evaluate different objectives. Two-stage ARO problems with column-wise uncertainties are considered in, e.g., Minoux (2011), Ardestani-Jaafari and Delage (2016a), Xu and Burer (2016).

Example 2 (removing redundant constraints for lot-sizing on a network). Let us again consider (P) in Example 1. The uncertain demand $\boldsymbol{z}$ is assumed to be in a budget uncertainty set:

$$
\mathcal{W} = \left\{\boldsymbol{z}\in\mathbb{R}_+^N \;\mid\; \boldsymbol{z}\leq\boldsymbol{20}, \boldsymbol{1}'\boldsymbol{z}\leq 20\sqrt{N}\right\}. \tag{20}
$$

Table 1 Removing redundant constraints for lot-sizing on a network. Here, "–" stands for not applicable, and "*" 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 Fourier-Motzkin elimination and RCI alternately; 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.

	#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	43594	*	*	–	–	–
	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	1096	12521	*	*	*	*
	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	31560	*	*
	Before	111	106	104	102	101	102	125	398	1359	*
	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	*

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., $X = \left\{ \boldsymbol{x} \in \mathbb{R}_+^{N_1} \mid \boldsymbol{x} \leq \mathbf{20} \right\}$. 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 redundant constraints identification (RCI) procedure, we repeatedly perform the following procedure: after eliminating an adjustable variable via Algorithm 1, we solve (19) for each constraint, and 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 2015) on an Intel i5-2400 3.10GHz Windows 7 computer with 4GB 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 0, if #Elim. $\leq N$. This is because we first eliminate the adjustable variables that have transport costs $t_{ii} = 0$, $i \in [N]$.
$\square$

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. (2017) that by adopting the lifted conic representable ambiguity set of Wiesemann et al. (2014), Problem (1) is also capable of modeling an adjustable distributionally robust optimization (ADRO) problem,

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

where $\tilde{\boldsymbol{z}}$ 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

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

with parameters $\boldsymbol{G} \in \mathbb{R}^{L_1 \times I_1}$ 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. (2017), Wiesemann et al. (2014) as part of $\tilde{\boldsymbol{z}}$ and we refer interested readers to their papers regarding the modeling capabilities of such an ambiguity set. Under the Slater's condition, i.e., the relative interior of $\{\boldsymbol{z} \in \mathcal{W} \;:\; \boldsymbol{G}\boldsymbol{z} \leq \boldsymbol{\mu}\}$ is non-empty, by introducing new here-and-now decision variables r and $\boldsymbol{s}$, Bertsimas et al. (2017) reformulate (21) into the following equivalent two-stage ARO problem,

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

where

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

We can now apply our approach to solve the above problem. In §7.2, we show that our approach significantly improves the obtained solutions in Bertsimas et al. (2017).

6. Generalization to multistage problems

The order of events in multistage ARO problems is as follows: The here-and-now decisions $\boldsymbol{x}$ are made before any uncertainty is realized, and then the uncertain parameters $\boldsymbol{z}_{\mathcal{S}^i}$ 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 $\boldsymbol{z}_{\mathcal{S}^i}$, but with no other knowledge of the uncertain parameters $\boldsymbol{z}_{\setminus\mathcal{S}^i}$ 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 $\mathcal{S}^i \cap \mathcal{S}^j = \emptyset$.

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:

$$\mathcal{X} = \left\{ \boldsymbol{x} \in X \mid \exists y_i \in \mathcal{R}^{|\mathcal{S}^i|,1}, \forall i \in [N_2] : \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{d}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{W} \right\}, \tag{22}$$

where y_i is the i-th element of $\boldsymbol{y}$, and $\boldsymbol{y}(\boldsymbol{z}) = [y_1(\boldsymbol{z}_{\mathcal{S}^1}), \cdots, y_{N_2}(\boldsymbol{z}_{\mathcal{S}^{N_2}})]' \in \mathcal{R}^{|\mathcal{S}^{N_2}|,N_2}$. While this process of decision making across stage is simple enough to state, modeling these *nonanticipativity restrictions*, i.e., 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 $l = \arg\max_{i \in [N_2]} |\mathcal{S}^i|$. Similarly as in Step 2 of Algorithm 1, we have the following constraints: there exists $y_i \in \mathcal{R}^{|\mathcal{S}^i|,1}$ for all $i \in [N_2] \setminus \{l\}$,

$$f_j(\overline{\boldsymbol{z}}) + \boldsymbol{g}_j'(\overline{\boldsymbol{z}})\boldsymbol{x} + \boldsymbol{h}_j'\boldsymbol{y}_{\setminus\{l\}}(\boldsymbol{z}) \geq f_i(\boldsymbol{z}) + \boldsymbol{g}_i'(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{h}_i'\boldsymbol{y}_{\setminus\{l\}}(\boldsymbol{z}) \quad \forall (\boldsymbol{z}, \overline{\boldsymbol{z}}) \in \overline{\mathcal{W}} \quad \text{if } b_{il} > 0 \text{ and } b_{jl} < 0, \tag{23}$$

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

where $\overline{\mathcal{W}} = \{(\boldsymbol{z}, \overline{\boldsymbol{z}}) \in \mathbb{R}^{2I_1} \mid \boldsymbol{z} \in \mathcal{W},\ \overline{\boldsymbol{z}} \in \mathcal{W},\ \boldsymbol{z}_{\mathcal{S}^l} = \overline{\boldsymbol{z}}_{\mathcal{S}^l}\}$ is an *augmented uncertainty set*. Due to the nonanticipitivity restrictions, the adjustable variable y_l only depends on $\boldsymbol{z}_{\mathcal{S}^l}$. The augmented uncertainty set $\overline{\mathcal{W}}$ enforces the constraints containing y_l to share the same information $\boldsymbol{z}_{\mathcal{S}^l}$, but the unrevealed $\boldsymbol{z}_{\setminus\mathcal{S}^l}$ are not necessarily the same across constraints. One simple yet crucial observation is that the nesting condition implies $\boldsymbol{y}_{\setminus\{l\}}(\overline{\boldsymbol{z}}) = \boldsymbol{y}_{\setminus\{l\}}(\boldsymbol{z})$ for all $(\boldsymbol{z}, \overline{\boldsymbol{z}}) \in \overline{\mathcal{W}}$. Hence, on the left hand side of the inequalities (23), we have $\boldsymbol{y}_{\setminus\{l\}}(\boldsymbol{z})$ instead of $\boldsymbol{y}_{\setminus\{l\}}(\overline{\boldsymbol{z}})$. 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. Firstly, we further investigate the lot-sizing problem discussed in Example 1 & 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:

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

with the dual uncertainty set:

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

Note that due to the existence of $\sum_{i\in[N]} \omega_i x_i$ in the first constraint of (D), the uncertainties are not column-wise. The RCI procedure proposed in §4 does not detect any redundant constraint. Hence, we only apply Algorithm 1 (without RCI) for (D). Here, the dimensions of adjustable variables in primal and dual formulations are significantly different, i.e., 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.

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 solve the reformulation with LDRs leads to better solutions.

Note that the optimal objective values (OPT) used in Table 2 are computed by enumerating all the vertices of the budget uncertainty set (20). For $N \leq 10$, the problems can be solved in 5 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 §2), such that we eliminate as many adjustable variables as possible while keeping the problem size at its minimal size.

Table 2 Lot-sizing on a Network for $N \in \{5, 10\}$. 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 RCI procedure; FME denotes the number of constraints from Algorithm 1; Gap% denotes the average optimality gap (in %) of 10 replications, i.e., for a candidate solution *sol.*, the gap is $\frac{sol.-OPT}{OPT}$, where OPT denotes the optimal objective value; Time records time (in seconds) needed to solve the corresponding optimization problem; TTime reports the total time (in seconds) needed to remove the redundant constraints and solve the optimization problem.

N=5	P	#Elim.	1	11	15	19	22	25	–
		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	–
	D	#Elim.	1	2	3	4	5	6	–
		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	–
N=10	P	#Elim.	1	12	17	19	21	22	100
		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	1639.8	*
	D	#Elim.	1	5	7	8	9	10	11
		FME	21	43	135	261	515	1025	149424
		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	*

Table 3 Lot-sizing on a Network for $N \in \{15, 20, 30\}$. Here, "$*$" indicates the average computation time exceeded the 10 min. 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 %) of the approximated solution via LDRs (without constraint elimination) of 10 replications , i.e., for a candidate solution *sol.*, the Red.% is $\frac{sol.-LDR}{LDR}$; Time records the time (in seconds) needed to solve the corresponding optimization problem.

	#Elim.	1	2	3	4	5	6	7	8	9	10	11
N=15	FME	31	31	33	39	53	83	145	271	525	1035	2057
	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	1045	2067
	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	1065	2087
	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	*

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), Bertsimas et al. (2017).

The problem setting here is adopted from Bertsimas et al. (2017). 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 to represent the inter-arrival time between patient j to the adjacent patient $j+1$ for $j \in [N-1]$ and 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 $i,\ i \in [N], i \geq 2$ 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:

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

where

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

Note that the introduction of the axillary random variable $\tilde{\boldsymbol{u}}$ in the ambiguity set is first introduced in Wiesemann et al. (2014) to obtain tractable formulations. Subsequently, Bertsimas et al. (2017) 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 $\boldsymbol{x}$ can be determined by solving the following two-stage adjustable distributionally robust optimization problem:

$$\begin{array}{lll} \min\limits_{\boldsymbol{x}, \boldsymbol{y}} & \sup\limits_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\sum_{i\in[N]} y_i(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) + \gamma y_{N+1}(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \right) & \\ \text{s.t.} & y_i(\boldsymbol{z}, \boldsymbol{u}) - y_{i-1}(\boldsymbol{z}, \boldsymbol{u}) + x_{i-1} \geq z_{i-1} & \forall (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{W} \quad \forall i \in \{2, \ldots, N+1\} \\ & \boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) \geq \boldsymbol{0} & \forall (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{W} \\ & \sum_{i\in[N]} x_i \leq T & \\ & \boldsymbol{x} \in \mathbb{R}^N_+, \boldsymbol{y} \in \mathcal{R}^{I_1+I_2, N+1}, & \end{array} \tag{25}$$

where y_i denotes the waiting time of patient $i, i \in [N]$, and y_{N+1} represents the overtime of the physician. Since $\mathcal{W}$ is clearly not polyhedral, the reformulation technique of Bertsimas and de

Ruiter (2016) cannot be applied here. As in Bertsimas et al. (2017), we use ROC to formulate the problem via LDRs, where the adjustable variables $\boldsymbol{y}$ are affinely in both $\boldsymbol{z}$ and $\boldsymbol{u}$, and solve it using CPLEX 12.6. ROC is a software package that is developed in C++ programming language and we refer readers to `http://www.meilinzhang.com/software` for more information.

Numerical study

The numerical settings of our computational experiments are similar to Bertsimas et al. (2017). 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 to 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. (2017) 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 2 seconds. Lastly, same as for the primal formulation of the lot-sizing problem, we observe no clear effect on the obtained results if different eliminating sequences are 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. (2017) for ADRO problems. Via numerical

Table 4 Appointment scheduling for $N=8$. Here, "$*$" 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; Min. Gap%, Max. Gap% and Gap% records the minimum, maximum and average optimality gap (in %) of 10 replications, respectively, i.e., for a candidate solution *sol.*, the gap is $\frac{sol.-OPT}{OPT}$, where OPT denotes the optimal objective value. All numbers reported in the last four rows are the average of 10 replications.

# Elim.	0	1	2	3	4	5	6	7	8	9
FME	18	17	17	20	37	132	731	5050	40329	*
Before	18	17	17	20	29	52	107	234	521	1152
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

experiments, we show that for small-size ARO problems our approach finds the optimal solution, and for moderate to large-size 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, e.g., 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 non-fixed 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 computing polytopic projections and identifying redundant constraints in linear programming problems. For instance, Huynh et al. (1992) discusses the efficiency of three alternative procedures for computing polytopic projections, and introduces a new RCI method; Paulraj and Sumathi (2010) compares 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.

A. Proof of the dual formulation (10) for Problem (1)

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

$$\min_{\boldsymbol{x}\in X}\max_{\boldsymbol{z}\in\mathcal{W}}\min_{\boldsymbol{y}}\left\{\ \boldsymbol{c}'\boldsymbol{x}\ \mid\ \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}\geq\boldsymbol{d}(\boldsymbol{z})\right\}.$$

Due to strong duality, we obtain the following reformulation by dualizing $\boldsymbol{y}$:

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

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

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

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

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

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

References

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

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

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

Ben-Ameur, W., G. Wang, A. Ouorou, M. Zotkiewicz (2016) Multipolar Robust Optimization. Available at: `arxiv.org/pdf/1604.01813.pdf`.

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

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

Ben-Tal, A., O. El Housni, V. Goyal (2016) A tractable approach for designing piecewise affine policies in dynamic robust optimization. Available at: `optimization-online.org/DB_FILE/2016/07/5557.pdf`.

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

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

Ben-Tal, A., A. Nemirovski (1999) Robust solutions of uncertain linear programs. *Operations Research Letters* 25:1–13.

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

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

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

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

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

Bertsimas, D., I. Dunning, M. Lubin (2015) Reformulation versus cutting-planes for robust optimization. *Computational Management Science* 13(2):195–217.

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

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

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

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

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

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

Bertsimas, D., M. Sim, M. Zhang (2017) A practically efficient approach for solving adaptive distributionally robust linear optimization problems. *Management Science*, to appear (available at: `optimization-online.org/DB_FILE/2016/03/5353.pdf`).

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

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

Breton, M., S. El Hachem (1995) Algorithms for the solution of stochastic dynamic minimax problems. *Computational Optimization and Applications* 4:317–345.

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

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

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

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

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

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

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

Dupacova, J. (1987) The minimax approach to stochastic programming and an illustrative application. *Stochastics* 20(1):73–88.

El Ghaoui, L., H. Lebret (1997) Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications* 18(4):1035–1064.

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

Hadjiyiannis, M., P. Goulart, D. Kuhn (2011) A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, Orlando, USA (IEEE, Piscataway, NJ), 7386–7391.

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

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

J. Fourier (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.

Goh, J., M. Sim (2009) Robust optimization made easy with ROME. *Operations Research* 59(4):973–985.

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

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

Gorissen, B., D. den Hertog (2013) Robust counterparts of inequalities containing sums of maxima of linear functions. *European Journal of Operational Research* 227(1):30–43.

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

Kali, P., S. Wallace (1995) *Stochastic Programming*. John Wiley & Sons.

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

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

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

M. Minoux (2011) On 2-stage robust LP with RHS uncertainty: complexity results and applications. *Journal of Global Optimization* 49:521—537.

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

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

Paulraj, S., P. Sumathi (2010) A comparative study of redundant constraints identification methods in linear programming Problems. *Mathematical Problems in Engineering*, vol. 2010.

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

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

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

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

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

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

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

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

Xu, G., S. Burer (2016) A copositive approach for two-stage adjustable robust optimization with uncertain right-hand sides. Available at: `arxiv.org/pdf/1609.07402v1.pdf`.

Xu, H., S. Mannor (2012) Distributionally robust Markov decision processes. *Mathematics of Operations Research* 37(2):288–300.

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

Zhen, J. and D. den Hertog (2017) Computing the maximum volume inscribed ellipsoid of a polytopic projection. *INFORMS Journal on Computing*, 30(1):31-42.

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