Adjustable Robust Optimization via Fourier-Motzkin Elimination

Jianzhe Zhen, Dick den Hertog
Department of Econometrics and Operations Research, Tilburg University

Melvyn Sim
NUS Business School, National University of Singapore

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 broader 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 used to reformulate ARO problems. This generic reformulation technique, contrasts with the classical approximation scheme via linear decision rules, enables us to solve adjustable optimization problems to optimality. We show via numerical experiments that, under limited computational resources, for small-size ARO problems our novel approach finds the optimal solution, and for moderate or large-size instances, we successively improve the solutions from linear decision rule type of approximations.

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 a dominant approach to address decision making under uncertainty. In classical 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.

In the seminal work Ben-Tal et al. (2004) extends classical robust optimization to encompass adjustable decisions. In these adjustable robust optimization (ARO) problems some of the decisions
have to be made here-and-now before the realization of the uncertain parameter is known. The other decisions are of a wait-and-see type, which are chosen after the uncertain parameter is realized. ARO problems are in general computationally intractable. To circumvent the intractability, Ben-Tal et al. (2004) restrict the adjustable decisions to be affinely dependent of the uncertain parameters, an approach known as linear decision rule (LDR) approximation.

Bertsimas et al. (2010), Bertsimas and Goyal (2012) establish the optimality of LDR approximation in some important classes of ARO problems. Chen and Zhang (2009) further improve the LDR approximation by extending the affine dependency to the auxiliary variables that are used in describing the uncertainty set. Henceforth, variants of piecewise affine decision rule approximation have been proposed to improve the approximation while maintaining the tractability of the adjustable distributionally optimization (ADRO) models. Such approaches include the deflected and segregated LDR approximation of Chen et al. (2008), the truncated LDR approximation of See and Sim (2009), and the bideflected and (generalized) segregated LDR approximation of Goh and Sim (2010). In fact, LDR approximation was 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 LDR approximation for solving multistage 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. (2014)).

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, 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 EI 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 single stage convex optimization problems, Wiesemann et al. (2014) propose a broad class of ambiguity set where the family of probability distributions are characterized by conic representable expectation constraints and nested conic representable confidence sets. Chen et al. (2007) adopt LDR approximation to provide tractable formulations for solving adjustable distributionally robust optimization (ADRO) problems. Bertsimas et al. (2016) incorporate the primary and auxiliary random variables of the lifted ambiguity set in the enhanced LDR approximation for ADRO problems, which lead to significantly better solutions than conventional LDR approximation.

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 rule approximations, finite adaptabilities, to improve the quality of obtained solutions (if not optimal). A description of FME procedure can be found in Fourier (1826) and Motzkin (1936). Via FME, we reformulate ARO problems into their equivalent counterparts with 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. Zhen and den Hertog (2015a) and Zhen and den Hertog (2015b) apply FME to solve a specific two-stage ARO problem, i.e., computing the maximum volume inscribed ellipsoid of a polytopic projection, but soon abandon this technique due to intractability. 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 (some of) 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.

Through the lens of FME, we investigate two-stage ARO problems theoretically, and prove that there exist convex piecewise affine functions, and concave piecewise affine functions that are ODRs for the adjustable variables. While eliminating adjustable variables in ARO problems
via FME, the structure of the uncertainty set is ignored. For a polyhedral uncertainty set, the equivalent dual formulation of Bertsimas and de Ruiter (2016) associates the constraints with the uncertainty set such that the structure of the uncertainty set in the primal formulation becomes part of the constraints in the dual formulation. By investigating dual formulation via FME, we characterize the structures of the ODRs for a broader class of ARO problems: a.) for two-stage ARO problems with simplex uncertainty sets, we show that there exist LDRs that are ODRs for the adjustable variables (which generalizes the result of Bertsimas and Goyal (2012), where authors only considers right-hand side uncertainties); b.) for two-stage ARO problems with box uncertainty sets, there exist two-piecewise affine functions that are ODRs for the adjustable variables in the dual formulation, and these problems can be casted as a sum-of-max problem. We also note that, despite the equivalence of primal and dual formulations, they may have significantly different number of adjustable variables. We evaluate the efficiency of our approach on both primal and dual formulations numerically. Since FME is very sensitive to the number of adjustable variables, our approach is particularly effective for the formulations with less adjustable variables. By using our FME approach, we extend the approach of Bertsimas et al. (2016) for ADRO problems. Via numerical experiments, we show that our approach improves the obtained solutions in Bertsimas et al. (2016).

Our main contributions are as follows:

1. We present a high level generic approach for ARO problems, which can easily incorporate with existing methods, e.g., LDR approximations and its variants, finite adaptabilities.
2. We investigate ARO problems via FME, which allows us to characterize the structures of the ODRs for a broader class of ARO problems.
3. We adapt an LP-based RCI procedure for ARO problems, which greatly reduces the number of constraints, and improves the efficiency of our proposed approach.
4. We show that our FME approach can be used to extend the approach of Bertsimas et al. (2016) for ADRO problems.
5. Via numerical experiments, we show that our approach can significantly improve the solutions from LDR type of approximations. Moreover, our reformulation technique via FME and RCI is independent from the objective function and uncertainty set of ARO problems and depends only on the constraints. Therefore, for a ARO problem, the reformulation can be pre-computed offline and used to evaluate different objectives and uncertainty sets efficiently.

This paper is organized as follows: In §2, we introduce FME for two-stage ARO problems. In §3, we investigate the primal and dual formulations of ARO problems with polyhedral uncertainty sets, and present some new results on the structures of the ODRs for several classes of problems. In §4, we propose an LP-based RCI procedure to effectively eliminate the redundant constraints.
§5 uses our FME approach to extend the approach of Bertsimas et al. (2016) for ADRO problems. We evaluate our approach numerically via lot-sizing on a network and appointment scheduling problems in §6. §7 presents conclusions and future research.

Notations. We use $[N], N \in \mathbb{N}$ to denote the set of running indices, \{1, $\ldots$, $N$\}. We generally use bold faced characters such as $x \in \mathbb{R}^N$ and $A \in \mathbb{R}^{M \times N}$ to represent vectors and matrices, respectively, and $x_S \in \mathbb{R}^{|S|}$ to denote a vector that contains a subset $S \subseteq [N]$ of components in $x$, e.g., $x_i$ or $x_i \in \mathbb{R}$ denotes the $i$-th element of $x$. We use $(x)^+$ to denote $\max\{x, 0\}$. Special vectors include $0$, $1$ and $e_i$ 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 $\tilde{z} \in \mathbb{R}^I$ to represent an $I$ dimensional random variable and it can be associated with a probability distribution $P \in \mathcal{P}_0(\mathbb{R}^I)$, where $\mathcal{P}_0(\mathbb{R}^I)$ represents the set of all probability distributions on $\mathbb{R}^I$. We denote $\mathbb{E}_P(\cdot)$ as the expectation over the probability distribution $P$. For a set $\mathcal{W} \subseteq \mathbb{R}^I$, $P(\tilde{z} \in \mathcal{W})$ represents the probability of $\tilde{z}$ being in the set $\mathcal{W}$ evaluated on the distribution $P$.

2. Two-stage adjustable robust optimization via Fourier-Motzkin elimination

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

$$\min_{x \in X} c'x,$$

where the feasible set $X$ is the set of all feasible here-and-now decisions:

$$X = \{ x \in X \mid \exists y \in \mathcal{R}^{I_1 \times N_2} : A(z)x + By(z) \geq d(z) \; \forall z \in W \},$$

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

$$A(z) = A^0 + \sum_{k \in [I_1]} A^k z_k, \quad d(z) = d^0 + \sum_{k \in [I_1]} d^k z_k,$$
with $A^0, A^1, ..., A^I \in \mathbb{R}^{M \times N_1}$ and $d^0, d^1, ..., d^I \in \mathbb{R}^M$. The matrix $B \in \mathbb{R}^{M \times N_2}$, also known in stochastic programming as the recourse matrix, is constant, which correspond 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. (2016) shows that it is capable of modeling adjustable distributionally robust optimization (ADRO) problems. In §5, we briefly discuss and extend their approach for ADRO problems. Problem (1) is generally intractable, even if there are only right hand side uncertainties (see Minoux (2011)), because the adjustable variables $y$ are decision rules instead of finite vectors of decision variables.

We propose to derive an equivalent representation of $X$ by eliminating the adjustable variables $y$ via Fourier-Motzkin elimination (FME). Algorithm 1 describes the FME procedure to eliminate an adjustable variable $y_l$, where $l \in [N_2]$. This algorithm is adapted from (Bertsimas and Tsitsiklis 1997, page 72) for polyhedral projections.

\section*{Algorithm 1 Fourier-Motzkin Elimination for two-stage problems.}

1. Rewrite each constraint in $X$ in the form: there exists $y \in \mathbb{R}^{I_1 \times N_2}$,
   \[
   b_{il}y_l(z) \geq d_i(z) - \sum_{j \in [N_1]} a_{ij}(z)x_j - \sum_{j \in [N_2] \setminus \{l\}} b_{ij}y_j(z) \quad \forall z \in W \quad \forall i \in [M];
   \]
   if $b_{il} \neq 0$, divide both sides by $b_{il}$. We obtain an equivalent representation of $X$ involving the following constraints: there exists $y \in \mathbb{R}^{I_1 \times N_2}$,
   \[
   y_l(z) \geq f_i(z) + g_i(z)x + h_i^lz_{(\{l\})}(z) \quad \forall z \in W \quad \text{if } b_{il} > 0,
   \]
   \[
   f_j(z) + g_j^z(z)x + h_j^y_{(\{l\})}(z) \geq y_l(z) \quad \forall z \in W \quad \text{if } b_{jl} < 0,
   \]
   \[
   0 \geq f_k(z) + g_k^z(z)x + h_k^y_{(\{l\})}(z) \quad \forall z \in W \quad \text{if } b_{kl} = 0.
   \]
   Here, each $h_i, h_j, h_k$ is a vector in $\mathbb{R}^{N_2-1}$, for a given $z$, each $f_l, f_j, f_k$ is a scalar, and each $g_l, g_j, g_k$ is a vector in $\mathbb{R}^{N_1}$.

2. Let $X_{\setminus \{l\}}$ be the feasible set after the adjustable variable $y_l$ is eliminated, and it is defined by the following constraints: there exists $y_{\setminus \{l\}} \in \mathbb{R}^{I_1 \times N_2-1}$,
   \[
   f_j(z) + g_j^z(z)x + h_j^y_{(\{l\})}(z) \geq f_i(z) + g_i^z(z)x + h_i^y_{(\{l\})}(z) \quad \forall z \in W \quad \text{if } b_{jl} < 0 \text{ and } b_{il} > 0
   \]
   \[
   0 \geq f_k(z) + g_k^z(z)x + h_k^y_{(\{l\})}(z) \quad \forall z \in W \quad \text{if } b_{kl} = 0.
   \]

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 $O(M^{2N_2})$, which is an unfortunate inheritance of FME.
**Theorem 1.** $\mathcal{X} = \mathcal{X}_{\{t\}}$.

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

We will now prove $\mathcal{X}_{\{t\}} \subset \mathcal{X}$. Let $x \in \mathcal{X}_{\{t\}}$. It follows from (6) that, there exists some $y_{\{t\}}(z)$,

$$
\min_{\{j | b_{ji} < 0\}} f_j(z) + g'_j(z)x + h'_jy_{\{t\}}(z) \geq \max_{\{i | b_{it} > 0\}} f_i(z) + g'_i(z)x + h'_iy_{\{t\}}(z), \quad \forall z \in \mathcal{W}.
$$

Let

$$
y_i(z) = \theta \min_{\{j | b_{ji} < 0\}} \{ f_j(z) + g'_j(z)x + h'_jy_{\{t\}}(z) \} + (1 - \theta) \max_{\{i | b_{it} > 0\}} \{ f_i(z) + g'_i(z)x + h'_iy_{\{t\}}(z) \}
$$

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

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

$$
\min_{x \in \mathcal{X}_{\{N_2\}}} c'x = \min_{x \in \mathcal{X}_{\{N_2\}}} c'x.
$$

Since the uncertainties in $\mathcal{X}_{\{N_2\}}$ are constraint-wise, we can now employ the conventional techniques from robust optimization to derive a tractable reformulation of each constraint in $\mathcal{X}_{\{N_2\}}$. Problem (8) can be solved to optimality if the constraints in $\mathcal{X}_{\{N_2\}}$ are not too many.

**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 $z$ 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 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:

$$
\begin{align*}
\min_{x \in X, y_{ij}, \gamma} & \quad c'x + \gamma \\
\text{s.t.} & \quad \sum_{i,j \in [N]} t_{ij}y_{ij}(z) \leq \gamma, \quad \forall z \in \mathcal{W} \\
& \quad z_i - x_i \leq \sum_{j \in [N]} y_{ji}(z) - \sum_{j \in [N]} y_{ij}(z), \quad \forall z \in \mathcal{W}, \quad i \in [N] \\
& \quad y_{ij}(z) \geq 0, \quad y_{ij} \in \mathcal{R}^{N,1}, \quad \forall z \in \mathcal{W}, \quad i, j \in [N].
\end{align*}
$$

(P)
The cost to transport one unit of demand from location $i$ to $j$ is denoted as $t_{ij}$; if $i = j$, $t_{ij} = 0$, otherwise, $t_{ij} \geq 0$. 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:

$$\min_{x \in X, y} \quad c'x + \gamma$$
$$\text{s.t.} \quad t_{21}z_1 - t_{21}x_1 \leq \gamma \quad \forall z \in W$$
$$t_{12}z_2 - t_{12}x_2 \leq \gamma \quad \forall z \in W$$
$$z_1 + z_2 - x_1 - x_2 \leq 0 \quad \forall z \in W$$
$$(t_{12} + t_{21})(z_1 - x_1) + t_{12}(x_2 - z_2) \leq \gamma \quad \forall z \in W.$$  

Note that we omit $\gamma \geq 0$, because it is clearly a redundant constraint, which can be easily detected in the elimination procedure. This is a static robust optimization problem. We show in §6.1 that imposing linear decision rules on $y_{ij}$ can lead to a suboptimal solution, whereas this equivalent reformulation produces the optimal solution.

As a result of Algorithm 1, the constraints in $X_{\alpha | [N_2]}$ may be prohibitively many. Then, we can first (iteratively) eliminate a subset $S \subseteq [N_2]$ of the adjustable variables in $X$ till the size of resulting description $X_{\alpha | S}$ reaches the prescribed computational limit, and then impose some simple functions on the remaining $y_i$, for all $i \in [N_2] \setminus S$. For instance, by imposing LDRs on $y_i$ in $X_{\alpha | S}$, for all $i \in [N_2] \setminus S$, the feasible set becomes:

$$\hat{X}_{\alpha | S} = \{ x \in X \mid \exists y \in \mathcal{L}^{I_1, N_2} : G(z)x + Hy(z) \geq f(z) \quad \forall z \in \mathcal{W} \}$$

where $G(z)$ and $H$ are the resulting coefficient matrices of $x$ and $y$, respectively, $f(z)$ is the corresponding right-hand side vector after elimination,

$$\mathcal{L}^{I_1, N_2} = \left\{ y \in \mathcal{R}^{I_1, N_2} \mid \exists y^0, y' \in \mathcal{R}^{N_2}, \quad i \in [I_1] : y(z) = y^0 + \sum_{i \in [I_1]} y' z_i \right\},$$

and $y' \in \mathcal{R}^{N_2}, i \in [I_1]$, are now decision variables. Since $y \in \mathcal{L}^{I_1, N_2} \subset \mathcal{R}^{I_1, N_2}$, it follows that $\hat{X}_{\alpha | S} \subseteq X_{\alpha | S}$. Hence, $\hat{X}_{\alpha | S}$ is a conservative (inner) approximation of $X_{\alpha | S}$. The following theorem shows that, the more adjustable variables are eliminated, the tighter the LDR approximation becomes; if all the adjustable variables are eliminated, the set representation is exact, i.e., $\hat{X}_{\alpha | [N_2]} = X_{\alpha | [N_2]} = X$.

**Theorem 2.** $\hat{X} \subseteq \hat{X}_{\alpha | S_1} \subseteq \hat{X}_{\alpha | S_2} \subseteq X$, for all $S_1 \subseteq S_2 \subseteq [N_2]$.

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

$$X_{\alpha | S} = \{ x \in X \mid \exists y_{\alpha | S} \in \mathcal{R}^{I_1, N_2 - |S|} : G(z)x + Hy_{\alpha | S}(z) \geq f(z) \quad \forall z \in \mathcal{W} \}$$,
where $G(z)$ and $H$ are the resulting coefficient matrices of $x$ and $y$, respectively, and $f(z)$ is the corresponding right-hand side vector after elimination. From Theorem 1, we know $\mathcal{X}_S = \mathcal{X}$. By imposing LDRs to the remaining $y_i$, for all $i \in [N_2] \setminus S$, by definition, we have

$$\hat{X}_{S} = \{ x \in X \mid \exists y_{i \setminus S} \in L_{f_1^{-1}}^{N_2-|S|} : G(z)x + Hy_{i \setminus S}(z) \geq f(z) \ \forall z \in W \},$$

where $A, B$ and $d$ are the same as in (2). Hence, it follows that $\hat{X} \subseteq \hat{X}_S \subseteq X_{\setminus S} = X$. Now, suppose $S_1 \subseteq S_2 \subseteq [N_2]$, we have $\hat{X}_{S_1} \subseteq \hat{X}_{S_2} \subseteq X_{\setminus S_1} = X_{\setminus S_2} = X$. \hfill \Box

Theorem 2 still holds if other decision rules, e.g., static decision rules, quadratic decision rules, polynomial decision rules, are imposed on the remaining adjustable variables. Moreover, instead of decision rule approximations, one can also adopt finite adaptability approaches to solve the reformulated ARO problems after some (not all) adjustable variables are eliminated.

3. Optimality of decision rules: a primal-dual perspective

In this section, we investigate the primal and dual formulations of 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 convex piecewise affine functions, and concave piecewise affine functions that are ODRs for the adjustable variables $y$ in Problem (1).

**Proof.** Let us denote $x^*$ as the optimal here-and-now decisions, and eliminate all the adjustable variables except for $y_l$ in $\mathcal{X}$ defined in (2) via Algorithm 1. Let $S_l = [N_2] \setminus \{ l \}$. From Theorem 1, we know $\mathcal{X} = \mathcal{X}_{\setminus S_l}$. The adjustable variable $y_l$ is upper (lower) bounded by the finite number of minimum (maximum) of affine functions in $z$, i.e.,

$$\hat{f}_l(z) \leq y_l(z) \leq \hat{f}_l(z) \ \forall z \in W,$$

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

$$\hat{f}_l(z) \leq \hat{f}_l(z) \ \forall z \in W$$


must hold and hence \( y_l(z) = \hat{f}_l(z) \) and \( y_l(z) = \tilde{f}_l(z) \) would be ODRs for the adjustable variable \( y_l \) in Problem (1) for all \( l \in [N_2] \).

Note that in Theorem 3, we do not impose any structural assumption on the uncertainty set \( W \) in Problem (1). Hence, Theorem 3 holds for Problem (1) with a general uncertainty set. Ben-Tal et al. (2016) mentions that there exist piecewise affine functions that are ODRs for Problem (1) with right hand side uncertainties which reside in a polytope. Let us now consider the following class of problems for which we prove in Proposition 1 that two-piecewise affine functions (with a specific structure) are ODRs.

**Example 2 (Production Planning).** A factory has \( N_1 \) machines that produce \( M \) types of goods. A manager would like to make a production plan today to satisfy the demands for tomorrow. The uncertainties in the demands and during the manufacturing process are captured by affine functions \( d(z) \in \mathcal{L}_{I_1,M} \) and \( A(z) \in \mathcal{L}_{I_1,M \times N_1} \), respectively, i.e., the machine \( j \) can produce \( a_{ij}(z) \) unit of product \( i \) per hour for all \( i \in [M] \), \( j \in [N_1] \), where \( z \in W \). The hourly cost of having machine \( j \) turned on is \( c_j \). The machine \( j \) is scheduled to be in production for \( x_j \) hours today.

The availability/capacity of the machines is incorporated in \( X \), i.e., \( x \in X \). In case the demand of good \( i \) is not fulfilled from the production, the manager have to purchase \( y_i \) unit of product \( i \) from a competitor at a price \( v_i \geq 0 \) per unit to fulfill the excess demands. The matrix \( P \in \mathbb{R}_{+}^{K \times M} \) and vector \( q \in \mathbb{R}_{+}^{K} \) can model the budget constraints, e.g., the total shortage of products cannot exceed a prescribed limit. To minimize the cost while satisfying all the demand, the manager can solve the following two-stage ARO problem:

\[
\begin{align*}
\min_{x \in X, y} & \quad c'x + \max_{z \in W} v'y(z) \\
\text{s.t.} & \quad Py(z) \leq q \quad \forall z \in W \\
& \quad A(z)x + y(z) \geq d(z) \quad \forall z \in W \\
& \quad y(z) \geq 0, \quad y \in \mathcal{R}_{I_1,M} \quad \forall z \in W.
\end{align*}
\]

(9)

As mentioned in §2, the worst-case second stage cost \( \max_{z \in W} v'y(z) \) in the objective can be modelled by Problem (1) using an epigraph formulation.

**Proposition 1.** There exist two-piecewise affine functions in the form of \( (d_i(z) - a'_i(z)x)^+ \) that are ODRs for the adjustable variables \( y_i, i \in [N_2] \), in Problem (9).
Theorem 3. Let us consider the equivalent epigraph formulation of Problem (9):

\[ \min_{x \in X, y, \gamma} \mathbf{c}'x + \gamma \]
\[ \text{s.t. } \mathbf{v}'y(z) \leq \gamma \quad \forall z \in \mathcal{W} \]
\[ P y(z) \leq \mathbf{q} \quad \forall z \in \mathcal{W} \]
\[ A(z)x + y(z) \geq \mathbf{d}(z) \quad \forall z \in \mathcal{W} \]
\[ y(z) \geq 0, \quad y \in \mathcal{R}_{I1,M}^+ \quad \forall z \in \mathcal{W}. \]

(10)

Let us denote \((x^*, \gamma^*)\) as the optimal here-and-now decisions, and eliminate all the adjustable variables except for \(y_l, l \in [N_2]\), in (10) via Algorithm 1. Since \(P\) and \(v\) are nonnegative, the adjustable variable \(y_l\) is only lower bounded by a two-piecewise affine function in \(z\), i.e.,

\[ \max\{d_l(z) - a_l'(z)x, 0\} \leq y_l(z) \quad \forall z \in \mathcal{W}, \]

where \(a_l\) are the \(l\)-th row vectors of matrix \(A\), and \(d_l\) is the \(l\)-th component of \(d\) for \(l \in [M]\). Hence, if Problem (10) is feasible, \(y_l(z) = \max\{d_l(z) - a_l'(z)x, 0\}\) would be an ODR for the adjustable variable \(y_l\). Analogously, it follows that, there exist ODRs in the form of \(y_l(z) = \max\{d_l(z) - a_l'(z)x, 0\}\) for all \(l \in [N_2]\).  

**3.2. A Dual Perspective**

Given a polyhedral uncertainty set

\[ \mathcal{W}_{poly} = \{ z \in \mathbb{R}_+^{I1} \mid P'z \leq \mathbf{q} \}, \]

where \(P \in \mathbb{R}^{I1 \times K}\) and \(q \in \mathbb{R}^K\), Bertsimas and de Ruiter (2016) derive an equivalent dual formulation of Problem (1):

\[ \min_{x \in \mathcal{X}^D} \mathbf{c}'x, \]

(11)

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

\[ \mathcal{X}^D = \left\{ x \in \mathcal{X} \mid \exists \mathbf{\lambda} \in \mathcal{R}^{M,K}_+ : \begin{array}{l}
\omega'(A^dx - d) - q'\mathbf{\lambda}(\omega) \geq 0 \quad \forall \omega \in \mathcal{U} \\
(\mathbf{p}_i'\mathbf{\lambda}(\omega)) \geq (d^i - A^ix)'\omega \quad \forall \omega \in \mathcal{U}, \forall i \in [I_1] \\
\mathbf{\lambda}(\omega) \geq 0 \quad \forall \omega \in \mathcal{U} \end{array} \right\} \]

(12)

with the dual uncertainty set:

\[ \mathcal{U} = \{ \omega \in \mathbb{R}_+^M \mid B'\omega = 0 \}, \]

where \(p_i \in \mathbb{R}^{I_1}\) are the \(i\)-th row vectors of matrix \(P\) for \(i \in [I_1]\). In (Bertsimas and de Ruiter 2016, Theorem 2), authors 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.

One can apply Algorithm 1 to eliminate adjustable variables in the dual formulation (11). Note that the structure of the uncertainty set in the primal formulation (1) becomes part of the constraints in the dual formulation (11). Eliminating adjustable variables in the primal formulation is equivalent to reducing the number of constraints in the uncertainty set \( U \) of the dual formulation, and vice versa. From Theorem 3, there exist convex (and concave) piecewise affine functions that are ODRs for the adjustable variables \( \lambda \) in the dual formulation (11). Let us consider two special classes of \( \mathcal{W}_{\text{poly}} \), i.e., a standard simplex and a box.

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

**Proof.** Suppose \( z \) reside in a standard simplex:

\[
\mathcal{W}_{\text{simplex}} = \{ z \in \mathbb{R}_{+}^{I_1} \mid 1' z \leq 1 \}.
\]

From (12), we have the following reformulation:

\[
\lambda^D = \left\{ x \in X \mid \exists \lambda \in \mathcal{R}^{M,1} : \omega'(A^0 x - d^0) - \lambda(\omega) \geq 0 \quad \forall \omega \in \mathcal{U} \\
\lambda(\omega) \geq \max\{ (d^i - A_i x)' \omega, 0 \} \quad \forall i \in [I_1], \forall \omega \in \mathcal{U} \right\}.
\]

Observe that the dual adjustable variable \( \lambda(\omega) \) is feasible in \( \lambda^D \) if and only if

\[
\omega'(A^0 x - d^0) \geq \lambda(\omega) \geq \max\{ \max_{i \in [I_1]} \{(d^i - A_i x)' \omega\}, 0 \} \quad \forall \omega \in \mathcal{U}.
\]

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

**Theorem 5.** Suppose the uncertainty set \( \mathcal{W}_{\text{poly}} \) is a box. Then, there exist two-piecewise affine functions in the form of \( (d^i - A_i x)' \omega \) that are ODRs for the adjustable variables \( \lambda_i, i \in [I_1] \) in Problem (11).

**Proof.** Suppose \( z \) reside in a box:

\[
\mathcal{W}_{\text{box}} = \{ z \in \mathbb{R}_{+}^{I_1} \mid z \leq q \}.
\]
where $q \in \mathbb{R}_{\geq 0}^{l_1}$. From (12), we have the following reformulation:

$$\mathcal{X}^D = \left\{ x \in X \mid \exists \lambda \in \mathbb{R}^{l_1} : \begin{array}{c}
\omega'(A^0 x - d^0) - q^\top \lambda(\omega) \geq 0 \\
\lambda_i(\omega) \geq \max\{(d' - A' x)\omega, 0\} \quad \forall \omega \in \mathcal{U} \\
\forall \omega \in \mathcal{U}, \forall i \in [l_1] \end{array} \right\}. \quad (13)$$

After eliminating all but the last adjustable variable $l$ in $\mathcal{X}^D$ via Algorithm 1, $l \in [l_1]$, the dual adjustable variable $\lambda_l(\omega)$ is feasible in $\mathcal{X}^D$ if and only if

$$\omega'(A^0 x - d^0)/q_l - \sum_{i \in [l_1]\setminus[l]} q_i \max\{(d' - A' x)\omega, 0\}/q_l \geq \lambda_l(\omega) \geq \max\{(d' - A' x)\omega, 0\} \quad \forall \omega \in \mathcal{U}. \quad (14)$$

One can observe that $\lambda_l$ is upper bounded by a $2^{l_1-1}$-piecewise affine function, and lower bounded by $((d' - A' x)\omega)^+$. Hence, there exists an ODR in the form of $\lambda_l(\omega) = ((d' - A' x)\omega)^+$, i.e., a two-piecewise affine function. Analogously, it follows that, there exist ODRs in the form of $\lambda_l(\omega) = ((d' - A' x)\omega)^+$ for all $i \in [l_1]$. \hfill \Box

An immediate observation from Theorem 5 is that, if we eliminate all the adjustable variables in (13) via Algorithm 1, it results in a sum-of-max representation:

$$\mathcal{X}^D_{[l_1]} = \left\{ x \in X \mid \forall \omega \in \mathcal{U} : \omega'(A^0 x - d^0) \geq \sum_{i \in [l_1]} q_i \max\{(d' - A' x)\omega, 0\} \right\}. \quad (15)$$

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

## 4. Removing redundant constraints

It is well-known that Fourier-Motzkin elimination often leads to many redundant constraints. In this section, we present an 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 $x \in X$ and $y \in \mathbb{R}^{l_1.N_2}$ such that

$$a^i(\zeta)x + b^i(y(\zeta)) \geq d_i(\zeta) \quad \forall \zeta \in \mathcal{W}, \forall i \in [M]\setminus\{l\}, \quad (16)$$

then

$$a^i(z)x + b^i(y(z)) \geq d_i(z) \quad \forall z \in \mathcal{W}, \quad (17)$$

where $a_i$ and $b_i$ are the $i$-th row vectors of matrices $A$ and $B$, respectively, and $d_i$ is the $i$-th component of $d$ 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 $x$. The following redundant constraint identification (RCI) procedure 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

\[
Z_l^* = \min_{x, y, z} \quad a'_l(z) x + b'_l y(z) - d_l(z) \\
\text{s.t.} \quad a'_l(\zeta) x + b'_l y(\zeta) \geq d_l(\zeta) \quad \forall \zeta \in \mathcal{W}, \quad \forall i \in [M] \setminus \{l\} \\
x \in X, \ y \in \mathbb{R}^{I_1}, \ z \in \mathcal{W}
\]

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

Proof. Indeed if \( Z_l^* \geq 0 \), then for all \( x \in X \) and \( y \in \mathbb{R}^{I_1} \) that are feasible in (14), we also have

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

which implies feasibility in (15). Conversely, if \( Z_l^* < 0 \), from the optimum solution of Problem (16), there exists a solution \( x \in X \) and \( y \in \mathbb{R}^{I_1} \) that would be feasible in (14), but

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

which would be infeasible in (15). □

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 \( M_1 \) and \( M_2 \) be two disjoint subsets of \( [M] \) such that

\[
a_i(z) = a_i, \ b_i \neq 0 \quad \forall i \in M_1, \\
\]

\[
b_i = 0 \quad \forall i \in M_2.
\]

Then the l-th constraint, \( l \in M_1 \) in the feasible set (2) is redundant if the following optimization problem:

\[
Z_l^* = \min_{x, y, z} \quad a'_l x + b'_l y - d_l(z) \\
\text{s.t.} \quad a'_l(\zeta) x \geq d_l(\zeta) \quad \forall \zeta \in \mathcal{W}, \quad \forall i \in M_2 \\
a'_l x + b'_l y \geq d_l(z) \quad \forall i \in M_1 \setminus \{l\} \\
x \in X, \ y \in \mathbb{R}^{I_2}, \ z \in \mathcal{W}
\]

has a nonnegative optimal objective value, i.e., \( Z_l^* \geq 0 \).
Proof. Observe that for any $l \in M_1$,

\[
Z_l^* \geq \min_{x, y, z} \ a_i'x + b_i'y(z) - d_i(z) \\
\text{s.t.} \quad a_i'(\zeta)x \geq d_i(\zeta) \quad \forall \zeta \in W, \ \forall i \in M_2 \\
a_i'x + b_i'y(\zeta) \geq d_i(\zeta) \quad \forall \zeta \in W, \ \forall i \in M_1 \{l\} \\
z \in W, x \in X, y \in R^{I_1, N_2}
\]

\[
\geq \min_{x, y, z} \ a_i'x + b_i'y(z) - d_i(z) \\
\text{s.t.} \quad a_i'(\zeta)x \geq d_i(\zeta) \quad \forall \zeta \in W, \ \forall i \in M_2 \\
a_i'x + b_i'y(z) \geq d_i(z) \quad \forall i \in M_1 \{l\} \\
z \in W, x \in X, y \in R^{I_1, N_2}
\]

\[
= \min_{x, y, z} \ a_i'x + b_i'y - d_i(z) \\
\text{s.t.} \quad a_i'(\zeta)x \geq d_i(\zeta) \quad \forall \zeta \in W, \ \forall i \in M_2 \\
a_i'x + b_i'y \geq d_i(z) \quad \forall i \in M_1 \{l\} \\
z \in W, x \in X, y \in R^{N_2}
\]

\[
= Z_l^*.
\]

Hence, whenever $Z_l^* \geq 0$, we have $Z_l^* \geq 0$, implying the $l$-th constraint is redundant. \qed

Note that in Theorem 7, to avoid intractability, only a subset of constraints in the feasible set (2) is considered, i.e., $M_1 \cup M_2 \neq [M]$. We can extend the subset $M_1 \subseteq [M]$ if the uncertainties affecting the constraints in $M_1$ are column-wise. Specifically let $\{z^0, ..., z^{N_1}\}, z^j \in R^{I_1}, j \in [N_1] \cup \{0\}$ be a partition of the vector $z \in R^{I_1}$ into $N_1 + 1$ vectors (including empty ones) such that

\[
W = \{(z^0, ..., z^{N_1}) \mid z^j \in W_j, \forall j \in [N_1] \cup \{0\}\}.
\]

(18)

Note that if $z^j, j \in [N_1]$ are empty vectors, then we would have $W = W_0$. Let $S \subseteq [N_1]$ and $\bar{S} = [N_1] \setminus S$ such that $x_j \geq 0$ for all $j \in S$ is implied by the set $X$. We redefine the subset $M_1 \subseteq [M]$ such that for all $i \in M_1$, $b_i \neq 0$ and the functions $a_{ij} \in L^{I_1, 1}$ and $d_i \in L^{I_1, 1}$ are affine in $z^j$, for all $j \in [N_2] \cup \{0\}$, specifically,

\[
a_{ij}(z) = a_{ij}(z^j) \quad \forall j \in S \\
a_{ij}(z) = a_{ij} \quad \forall j \in \bar{S} \\
d_i(z) = d_i(z^0).
\]

Note that since $S$ or $z^j, j \in [N_1]$ can be empty sets, the conditions to select $M_1$ is more general than in Theorem 7. From Theorem 7, one can check whether the $l$-th inequality, $l \in M_1$ is redundant
by solving the following problem:

\[
Z_l^* = \min_{x \in \mathcal{X}, y, z} \sum_{j \in \mathcal{S}} a_{lj} (z^j) x_j + \sum_{j \in \mathcal{S}} a_{ij} x_j + b'_i y - d_i (z^0)
\]

s.t. \[a'_i(\zeta) x \geq d_i(\zeta) \quad \forall \zeta \in \mathcal{W}, \forall i \in \mathcal{M}_2\] (19)

\[
\sum_{j \in \mathcal{S}} a_{ij} (z^j) x_j + \sum_{j \in \mathcal{S}} a_{ij} x_j + b'_i y \geq d_i (z^0) \quad \forall i \in \mathcal{M}_1 \setminus \{l\}\]

\(z^l \in \mathcal{W}_j\) \quad \forall j \in \mathcal{S} \cup \{0\}.

The \(l\)-th inequality is redundant if the optimal objective value is nonnegative. Due to the presence of products of variables (e.g., \(z^j x_j\)), Problem (19) is nonconvex in \(x\) and \(z\). An equivalent convex representation of (19) can be obtained by substituting \(w_j = z^j x_j, j \in \mathcal{S}\),

\[
Z_l^* = \min_{x \in \mathcal{X}, y, z} \sum_{j \in \mathcal{S}} a_{lj} (w^j / x_j) x_j + \sum_{j \in \mathcal{S}} a_{ij} x_j + b'_i y - d_i (z^0)
\]

s.t. \[a'_i(\zeta) x \geq d_i(\zeta) \quad \forall \zeta \in \mathcal{W}, \forall i \in \mathcal{M}_2\] (20)

\[
\sum_{j \in \mathcal{S}} a_{ij} (w^j / x_j) x_j + \sum_{j \in \mathcal{S}} a_{ij} x_j + b'_i y \geq d_i (z^0) \quad \forall i \in \mathcal{M}_1 \setminus \{l\}\]

\((w^j, x_j) \in K_j\) \quad \forall j \in \mathcal{S}

\(z^0 \in \mathcal{W}_0\),

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

\[K_j = \text{cl} \left\{ (u, t) \in \mathbb{R}^{I_j+1} \mid u / t \in \mathcal{W}_j, \ t > 0 \right\} .\]

Hence, (20) 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 (2015b) apply this technique to derive a convex representation of the feasible set for systems of uncertain linear equations.

It can be observed that Algorithm 1 does not destroy the column-wise uncertainties. Two-stage ARO problems with column-wise uncertainties are considered in, e.g., Minoux (2011), Bertsimas and de Ruiter (2016), Bertsimas et al. (2016).

**Example 3 (removing redundant constraints for lot-sizing on a network).** Let us again consider (P) in Example 1. The uncertain demand \(z\) is assumed to be in a budget uncertainty set:

\[\mathcal{W} = \left\{ z \in \mathbb{R}^N \mid z \leq 20, 1^t z \leq 20 \sqrt{N} \right\} .\]

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
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.

<table>
<thead>
<tr>
<th>#Elim.</th>
<th>0</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>12</th>
<th>13</th>
<th>16</th>
<th>19</th>
<th>22</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>FME</td>
<td>13</td>
<td>10</td>
<td>21</td>
<td>126</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Before</td>
<td>13</td>
<td>10</td>
<td>14</td>
<td>17</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>After</td>
<td>13</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Time</td>
<td>0</td>
<td>0.8</td>
<td>1.7</td>
<td>2.9</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>FME</td>
<td>21</td>
<td>17</td>
<td>20</td>
<td>60</td>
<td>43594</td>
<td>*</td>
<td>*</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Before</td>
<td>21</td>
<td>17</td>
<td>18</td>
<td>26</td>
<td>75</td>
<td>92</td>
<td>102</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>After</td>
<td>21</td>
<td>17</td>
<td>18</td>
<td>23</td>
<td>31</td>
<td>33</td>
<td>36</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Time</td>
<td>0</td>
<td>0.6</td>
<td>1.9</td>
<td>3.6</td>
<td>10.2</td>
<td>13.5</td>
<td>24.3</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>FME</td>
<td>31</td>
<td>26</td>
<td>26</td>
<td>37</td>
<td>1096</td>
<td>12521</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Before</td>
<td>31</td>
<td>26</td>
<td>26</td>
<td>31</td>
<td>76</td>
<td>108</td>
<td>486</td>
<td>697</td>
<td>869</td>
<td>750</td>
</tr>
<tr>
<td>After</td>
<td>31</td>
<td>26</td>
<td>25</td>
<td>27</td>
<td>46</td>
<td>54</td>
<td>82</td>
<td>101</td>
<td>116</td>
<td>127</td>
</tr>
<tr>
<td>Time</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
<td>3.1</td>
<td>9.0</td>
<td>13.0</td>
<td>54.0</td>
<td>137.3</td>
<td>247.8</td>
</tr>
<tr>
<td>FME</td>
<td>111</td>
<td>106</td>
<td>104</td>
<td>102</td>
<td>102</td>
<td>104</td>
<td>104</td>
<td>165</td>
<td>31560</td>
<td>*</td>
</tr>
<tr>
<td>Before</td>
<td>111</td>
<td>106</td>
<td>104</td>
<td>104</td>
<td>104</td>
<td>104</td>
<td>102</td>
<td>125</td>
<td>398</td>
<td>1359</td>
</tr>
<tr>
<td>After</td>
<td>111</td>
<td>106</td>
<td>104</td>
<td>104</td>
<td>102</td>
<td>100</td>
<td>116</td>
<td>180</td>
<td>343</td>
<td>–</td>
</tr>
<tr>
<td>Time</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.6</td>
<td>3.7</td>
<td>4.7</td>
<td>24.3</td>
<td>624.3</td>
<td>*</td>
</tr>
</tbody>
</table>

cost per unit is \(c_i = 20\), \(i \in [N]\), and the capacity of each store is 20, i.e., \(X = \{x \in \mathbb{R}^N_+ \mid x \leq 20\}\). The numerical settings here are adopted from Bertsimas and de Ruiter (2016). In Table 1, we illustrate the effectiveness of our procedure introduced above. To utilize the effectiveness of redundant constraints identification (RCI) procedure, we repeatedly perform the following procedure: after eliminating an adjustable variable via Algorithm 1, we solve (20) 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 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 43594 constraints, and the total time needed for detecting and removing the redundant constraints thus far is 9.6 seconds. Note that Time is 0, if #Elim. \(\leq N\). This is because we first eliminate the adjustable variables that have transport costs \(t_{ii} = 0\), \(i \in [N]\).
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. (2016) 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,

$$\begin{align*}
\min_{x,y} & \quad c'x + \sup_{P \in F} E_P (v' y(\tilde{x})) \\
\text{s.t.} & \quad A(z)x + By(z) \geq d(z) \quad \forall z \in \mathcal{W} \\
& \quad x \in X, y \in \mathbb{R}^{I_1,N_2},
\end{align*}$$

(21)

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

$$F = \left\{ P \in \mathcal{P}_0(\mathbb{R}^{I_1}) \mid E_P(G \tilde{z}) \leq \mu, P(\tilde{z} \in \mathcal{W}) = 1 \right\},$$

with parameters \(G \in \mathbb{R}^{L_1 \times I_1}\) and \(\mu \in \mathbb{R}^{L_1}\). For convenience and without loss of generality, we have incorporated the auxiliary random variable defined in Bertsimas et al. (2016), Wiesemann et al. (2014) as part of \(\tilde{z}\) and we refer interested readers to their papers regarding the modeling capabilities of such an ambiguity set. Under a Slater condition, i.e., the relative interior of \(\{z \in \mathcal{W} : Gz \leq \mu\}\) is non-empty, by introducing new here-and-now decision variables \(r\) and \(s\), Bertsimas et al. (2016) reformulate (21) into the following two-stage ARO problem,

$$\begin{align*}
\min_{(x,r,s) \in \tilde{X}} & \quad c'x + r + s'\mu \\
\text{s.t.} & \quad \exists y \in \mathbb{R}^{I_1,N_2} : r + s'(Gz) \geq v' y(z) \quad \forall z \in \mathcal{W} \\
& \quad A(z)x + By(z) \geq d(z) \quad \forall z \in \mathcal{W} \\
& \quad x \in X, r \in \mathbb{R}, s \in \mathbb{R}^{I_1,N_2},
\end{align*}$$

We can now apply our approach to solve the above problem. In §6.2, we show that our approach significantly improves the obtained solutions in Bertsimas et al. (2016).

6. Numerical experiments

6.1. Lot-sizing on a network

Let us again consider \((P)\) in Example 1 with the same parameter setting as in Example 3. From (12), one can write the equivalent dual formulation:

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

(D)
with the dual uncertainty set:

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

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 LDR leads to better solutions.

**Table 2**  Lot-sizing on a Network for \( N \in \{5, 10\} \). We use \#Elim. to denote the number of eliminated adjustable variables; FME denotes the number of constraints from Algorithm 1; RCI is the number of resulting constraints from first applying Algorithm 1 and then RCI procedure; %Gap denotes the average optimality gap (in %) of 10 replications, i.e., for a candidate solution \( \text{sol.} \), the gap is \( \frac{\text{sol.} - \text{OPT}}{\text{OPT}} \); 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.

<table>
<thead>
<tr>
<th>N=5</th>
<th>#Elim.</th>
<th>RCI</th>
<th>%Gap</th>
<th>TTime</th>
<th>TTime</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>1</td>
<td>30</td>
<td>3.3</td>
<td>0.1</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>37</td>
<td>2.9</td>
<td>12.9</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>75</td>
<td>0.7</td>
<td>58.3</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>101</td>
<td>0.1</td>
<td>223.2</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>116</td>
<td>0</td>
<td>394.3</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>127</td>
<td>–</td>
<td>550.3</td>
<td>–</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>3</td>
<td>3.3</td>
<td>11.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>13</td>
<td>2.8</td>
<td>10.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>19</td>
<td>2.3</td>
<td>19.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>23</td>
<td>1</td>
<td>23.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>33</td>
<td>0.2</td>
<td>33.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>272</td>
<td>0</td>
<td>272.0</td>
<td>–</td>
</tr>
<tr>
<td>N=10</td>
<td>#Elim.</td>
<td>RCI</td>
<td>%Gap</td>
<td>Time</td>
<td>Ttime</td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>110</td>
<td>6.0</td>
<td>0.1</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>100</td>
<td>6.0</td>
<td>14.8</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>133</td>
<td>5.9</td>
<td>52.6</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>180</td>
<td>5.8</td>
<td>100.7</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>276</td>
<td>5.7</td>
<td>987.7</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>343</td>
<td>5.7</td>
<td>1639.8</td>
<td>–</td>
</tr>
<tr>
<td>D</td>
<td>#Elim.</td>
<td>RCI</td>
<td>%Gap</td>
<td>Time</td>
<td>Ttime</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>41</td>
<td>6</td>
<td>21.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>44</td>
<td>4.6</td>
<td>44.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>135</td>
<td>2.6</td>
<td>135.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>261</td>
<td>1.8</td>
<td>261.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>515</td>
<td>0.8</td>
<td>515.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1025</td>
<td>0.2</td>
<td>1025.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>149424</td>
<td>0.2</td>
<td>149424</td>
<td>–</td>
</tr>
</tbody>
</table>
Table 3: Lot-sizing on a Network for $N \in \{15, 20, 30\}$. Here, “*” the average computation time exceeded 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 LDR (without constraint elimination) of 10 replications, i.e., for a candidate solution $sol.$, the %Red. is $\frac{sol.-LDR}{LDR}$.

<table>
<thead>
<tr>
<th>$N=15$</th>
<th>$N=20$</th>
<th>$N=30$</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Elim.</td>
<td>FME</td>
<td>%Red.</td>
</tr>
<tr>
<td>1</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>-0.1</td>
</tr>
<tr>
<td>3</td>
<td>33</td>
<td>-0.4</td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>-0.7</td>
</tr>
<tr>
<td>5</td>
<td>53</td>
<td>-0.9</td>
</tr>
<tr>
<td>6</td>
<td>145</td>
<td>-1.6</td>
</tr>
<tr>
<td>7</td>
<td>271</td>
<td>-1.9</td>
</tr>
<tr>
<td>8</td>
<td>525</td>
<td>-2.2</td>
</tr>
<tr>
<td>9</td>
<td>1035</td>
<td>-2.8</td>
</tr>
<tr>
<td>10</td>
<td>2057</td>
<td>-3.4</td>
</tr>
<tr>
<td>11</td>
<td>1862</td>
<td>-3.9</td>
</tr>
</tbody>
</table>

Time records time (in seconds) needed to solve the corresponding optimization problem.

6.2. An application in medical appointment scheduling

The problem settings here are adopted from Bertsimas et al. (2016). 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.

To formulate the problem, 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:

$$\mathcal{F} = \left\{ P \in \mathcal{P}_0(\mathbb{R}^N \times \mathbb{R}^{N+1}) \mid \begin{array}{l}
\mathbb{E}_P(\tilde{z}) = \mu \\
\mathbb{E}_P(\tilde{u}_i) \leq \phi_i \\
\mathbb{P}(\tilde{z}, \tilde{u} \in \mathcal{W}) = 1 
\end{array} \forall i \in [N+1] \right\},$$

where

$$\mathcal{W} = \left\{ (z, u) \in \mathbb{R}^N \times \mathbb{R}^{N+1} \mid \begin{array}{l}
z \geq 0 \\
(z_i - \mu_i)^2 \leq u_i \\
\left( \sum_{i \in [N]} (z_i - \mu_i) \right)^2 \leq u_{N+1} 
\end{array} \forall i \in [N] \right\}.$$
Note that the introduction of the axillary random variable $\tilde{u}$ in the ambiguity set is first introduced in Wiesemann et al. (2014) to obtain tractable formulations. Subsequently, Bertsimas et al. (2016) 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 $\gamma$ per unit delay. The optimal arrival schedule $x$ can be determined by solving the following two-stage adjustable distributionally robust optimization problem:

$$\min_{x, y} \sup_{P \in \mathcal{P}} \left( \sum_{i \in [N]} y_i(\tilde{z}, \tilde{u}) + \gamma y_{N+1}(\tilde{z}, \tilde{u}) \right)$$

s.t. $y_i(z, u) - y_{i-1}(z, u) + x_{i-1} \geq z_{i-1}$ \quad $\forall (z, u) \in W$ \quad $\forall i \in \{2, \ldots, N+1\}$

$y(z, u) \geq 0$ \quad $\forall (z, u) \in W$

$$\sum_{i \in [N]} x_i \leq T$$

$x \in \mathbb{R}^N, y \in \mathbb{R}^{I_1 + I_2, N+1}$,

where the $y_i$ denotes the waiting time of patient $i, i \in [N]$, and $y_{N+1}$ represents the overtime of the physician. Since $W$ is clearly not polyhedral, the reformulation technique of Bertsimas and de Ruiter (2016) cannot be applied here. As in Bertsimas et al. (2016), we use ROC to formulate the problem via LDR approximation, where the adjustable variables $y$ are affinely in both $z$ and $u$, 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 Mak et al. (2014), Bertsimas et al. (2016). We have $N = 8$ jobs and the unit overtime cost is $\gamma = 2$. For each job $i \in [N]$, we randomly select $\mu_i$ based on uniform distribution over $[30, 60]$ and $\sigma_i = \mu_i \cdot \epsilon$ where $\epsilon$ 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 (22), in which 1 to 9 adjustable variables are eliminated, with 10 randomly generated uncertainty sets. As shown in Table 4, 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 (22). For the 10 randomly generated uncertainty sets, the average optimality gap of the solutions obtained in Bertsimas et al. (2016) 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.

### 7. Conclusions

We propose a generic FME approach for solving ARO problems with fixed recourse. 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. (2016) for ADRO problems. Via numerical 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 solutions obtained from LDR type of approximations.

On theoretical level, one immediate future research direction would be to extend our framework for two-stage ARO problems to the multistage case, and characterize the structures of the ODRs for the multistage problems.

On 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.
References


