Pareto Adaptive Robust Optimality via a Fourier-Motzkin Elimination Lens

D. Bertsimas* S.C.M. ten Eikelder† D. den Hertog‡ N. Trichakis*

December 9, 2020

We introduce the concept of Pareto Adaptive Robust Optimality (PARO) for linear Adaptive Robust Optimization (ARO) problems. A worst-case optimal solution pair of here-and-now decisions and wait-and-see decisions is PARO if it cannot be Pareto dominated by another solution, i.e., there does not exist another such pair that performs at least as good in all scenarios in the uncertainty set and strictly better in at least one scenario. We argue that, unlike PARO, extant solution approaches—including those that adopt Pareto Robust Optimality from static robust optimization—could fail in ARO and yield solutions that can be Pareto dominated. The latter could lead to inefficiencies and suboptimal performance in practice. We prove the existence of PARO solutions, and present particular approaches for finding and approximating such solutions. We present numerical results for an inventory management and a facility location problem that demonstrate the practical value of PARO solutions.

Our analysis of PARO relies on an application of Fourier-Motzkin Elimination as a proof technique. We demonstrate how this technique can be valuable in the analysis of ARO problems, besides PARO. In particular, we employ it to devise more concise and more insightful proofs of known results on (worst-case) optimality of decision rule structures.

Keywords: Robust optimization; adaptive robust optimization; Pareto optimality; Fourier-Motzkin Elimination; decision rules

1 Introduction

Robust Optimization (RO) is a widespread methodology for modeling decision-making problems under uncertainty that seeks to optimize worst-case performance (Bertsimas et al., 2011; Gabrel et al., 2014; Gorissen et al., 2015). In practice, RO problems usually admit multiple worst-case optimal solutions, the performance of which may differ substantially under non-worst-case uncertainty scenarios. Consequently, the choice of an optimal solution often has material impact on performance under real-world implementations. This important consideration, which was first brought forth by Iancu and Trichakis (2014), has been successfully tackled for static, single-stage (linear) RO problems. For the increasingly popular and broad class of dynamic, multi-stage Adaptive Robust Optimization (ARO) problems (Ben-Tal et al., 2004), however, there is no successful approach for choosing an optimal solution, and the purpose of this paper is to bridge this gap.

*Operations Research Center and Sloan School of Management, Massachusetts Institute of Technology, USA
†Department of Econometrics and Operations Research, Tilburg University, The Netherlands
‡Department of Operations Management, University of Amsterdam, The Netherlands

In particular, for static RO problems, Iancu and Trichakis (2014) proposed the choice of so-called Pareto Robustly Optimal (PRO) solutions. In general, PRO solutions unarguably dominate non-PRO solutions, because, by definition, the former guarantee that there do not exist other worst-case optimal solutions that perform at least as good as the current solution for all scenarios in the uncertainty set, while performing strictly better for at least one scenario.

Going beyond static RO problems, it is well understood that the choice of an optimal solution remains crucial for the broader class of multi-stage ARO problems. Similar to RO solutions, by following a worst-case philosophy and not considering performance across the entire spectrum of possible scenarios, ARO optimal solutions could lead to substantial performance losses. For example, see the work by De Ruiter et al. (2016), who numerically demonstrate existence of multiple worst-case optimal solutions for the classical multi-stage inventory-production model that was considered in Ben-Tal et al. (2004), and find them to differ considerably from each other in their non-worst-case performance.

For ARO problems, however, a solution approach that unarguably chooses "good solutions," similar to PRO solutions for static RO problems, has proved to be elusive thus far. Extant approaches have all attempted to simply apply the concept of PRO to ARO problems. Specifically, they advocate restricting attention to adaptive variables that depend affinely on the uncertain parameters; commonly referred to as Linear Decision Rules (LDRs). Restricting to LDRs reduces the problem to static RO, and enables the search for associated PRO solutions (Iancu and Trichakis, 2014; De Ruiter et al., 2016; Bertsimas et al., 2019). As we shall show, however, this indirect application of the PRO concept fails to produce solutions that cannot be dominated.

In this paper, we introduce and study the concept of *Pareto Adaptive Robustly Optimal (PARO)* solutions for linear ARO problems. Similar to PRO solutions for static RO problems, PARO solutions yield worst-case optimal performance and are not dominated by any other such solutions in non-worst-case scenarios. In other words, PARO solutions unarguably dominate non-PARO solutions, leading to improved performance in non-worst-case scenarios, while maintaining worst-case optimality. From a practical standpoint, this means that implementing PARO solutions can only yield performance benefits, without any associated drawbacks.

To introduce the PARO concept and highlight its practical importance, we provide an illustrative toy example. The example also serves two additional important purposes. First, it enables us to show in a simple setting how PARO solutions can dominate PRO solutions, as remarked above. Second, the example motivates the need for new analysis techniques for studying PARO.

Example 1. In treatment planning for radiation therapy, the goal is to deliver a curative amount of dose to the target volume (tumor tissue), while minimizing the dose to healthy tissues. Consider a simplified case with two target subvolumes. For subvolume $i \in \{1, 2\}$, the required dose level d_i depends on the radiation sensitivity of the tissue, which is unknown. Assume that, prior to treatment, the doses lie in

$$U = \{(d_1, d_2) \mid 50 \leq d_i \leq 60,\ i = 1, 2\}.$$

Mid-treatment, the required doses are ascertained via biomarker measurements.

Treatment doses are administered in two stages. The dose administered in the first stage, denoted by x, needs to be decided prior to treatment. The dose administered in the second stage, denoted by y, can be decided after the required doses have been ascertained, i.e., it can be adapted to uncertainty revelation. Both treatment doses are delivered homogeneously over both volumes in each stage. Dose in each stage is limited to the interval $[20, 40]$. The total dose administered is $x + y$, and the healthy tissue receives a fraction $\delta > 0$ of it. The Stage-1 dose x, and a decision rule $y(\cdot)$ for the adaptive Stage-2 dose can be chosen by solving:

$$
\begin{aligned}
\min_{x,y(\cdot)} \quad & \max_{(d_1,d_2)\in U} \delta(x + y(d_1,d_2)) && \text{(1a)}\\
\text{s.t.} \quad & x + y(d_1,d_2) \geq d_1, \quad \forall (d_1,d_2) \in U, && \text{(1b)}\\
& x + y(d_1,d_2) \geq d_2, \quad \forall (d_1,d_2) \in U, && \text{(1c)}\\
& 20 \leq x \leq 40, && \text{(1d)}\\
& 20 \leq y(d_1,d_2) \leq 40, \quad \forall (d_1,d_2) \in U. && \text{(1e)}
\end{aligned}
$$

Problem (1) is an ARO problem with constraintwise uncertainty, for which static decision rules are worst-case optimal (Ben-Tal et al., 2004). Plugging in $y(d_1,d_2) = y$ and solving the resulting static RO model yields a worst-case optimal objective value of 60δ, achieved by all (x,y) such that $x+y = 60$. For any such solution, the objective value remains 60δ in not only the worst-case scenario but in all scenarios. Hence, all these solutions are PRO, according to the definition of Iancu and Trichakis (2014). Consequently, the Stage-1 decisions that are PRO lie in the set:

$$X^{\text{PRO}} = \{x \mid 20 \leq x \leq 40\}.$$

Consider now the decision rule $y^*(d_1,d_2) = \max\{20, d_1 - x, d_2 - x\}$, which is feasible for all feasible x. Furthermore, this rule minimizes the objective for any fixed x, d_1 and d_2. Plugging this in yields

$$\min_{20\leq x\leq 40} \max_{(d_1,d_2)\in U} \delta \max\{20 + x, d_1, d_2\}.$$

For given (d_1,d_2) the objective value is at least $\delta \max\{d_1,d_2\}$, and this is achieved by all $x \leq 30$. Thus, it should be preferable to implement one of these solutions for the Stage-1 decision. In fact, these solutions, which we call PARO, cannot be dominated by other solutions. Notably, the set of PARO solutions

$$X^{\text{PARO}} = \{x \mid 20 \leq x \leq 30\},$$

is a strict subset of X^{PRO}. This implies that PARO solutions could dominate PRO solutions that are non-PARO. To exemplify, consider PARO solution $x^* = 25$ and PRO (non-PARO) solution $\hat{x} = 35$. If $(d_1,d_2) = (60,60)$, both solutions yield worst-case objective value 60δ. If $(d_1,d_2) = (50,55)$, $\hat{x}$ still yields objective value 60δ whereas x^* yields objective value 55δ. There is no scenario where $\hat{x}$ results in a strictly better objective value than x^*. Thus, the PRO solution $\hat{x}$ is dominated by the PARO solution x^*. ▲

Besides showing that PRO solutions could be dominated in ARO problems, Example 1 also provides intuition into how. In particular, what unlocks extra performance in ARO problems is the application of decision rules that are not merely worst-case optimal, but rather "Pareto optimal," i.e., they optimize performance over non-worst-case scenarios as well. Note, however, that although for worst-case optimality linear decision rules might suffice under special circumstances, for Pareto optimality non-linear rules appear to be more often necessary, as illustrated by the example.

The application of non-linear decision rules to study PARO solutions invalidates the techniques used in the analysis of Pareto efficiency in RO in the extant literature, which is solely focused on linear formulations. In other words, analysis of Pareto efficiency in ARO calls for a new line of attack, which brings us to another contribution we make. Specifically, to study PARO solutions, we rely heavily on Fourier-Motzkin Elimination (FME) as a proof technique. Through the lens of FME we consider optimality of decision rule structures, which then enables us to study PARO. Furthermore, we illustrate how this proof technique can be applied in ARO more generally, by providing more general and more insightful proofs of known results (not related to Pareto efficiency).

Findings and Contributions

Before we begin our analysis, we summarize the findings and the contributions of this paper. The treatment presented is restricted to two-stage ARO models that are linear in both decision variables and uncertain parameters. The restriction to two stages is primarily motivated by ease of exposition and notational simplicity.

1. *Concept of PARO Solutions.* In the context of linear ARO problems, we introduce the concept of Pareto Adaptive Robustly Optimal (PARO) solutions. PARO solutions have the property that no other solution and associated adaptive decision rule exist that dominate them, i.e., perform at least as good under any scenario, and perform strictly better under at least some scenario. As Example 1 above has already shown, in the context of ARO problems, PARO solutions can dominate other Pareto optimal solution concepts already proposed in the literature (Iancu and Trichakis, 2014). In practice, PARO solutions can only yield performance benefits compared with non-PARO solutions, as the latter lead to efficiency losses.

2. *Properties of PARO Solutions.* We derive several properties of PARO solutions. Among them, we show that first-stage PARO solutions exist for any two-stage ARO problem with a compact feasible region. Furthermore, we prove that for any two-stage ARO problem there exists a piecewise linear (PWL) decision rule that is PARO. To arrive at these results, our analysis relies on FME.

3. *Finding PARO Solutions and their Practical Value.* We present several approaches to find and/or approximate PARO solutions in practice, amongst others using techniques based on FME. We also conduct numerical studies for an inventory management example and a facility location example. The results reveal that (approximate) PARO solutions can yield substantially better performance in non-worst-case scenarios than worst-case optimal and PRO solutions, thus demonstrating the practical value of the proposed methodology.

4. *FME as a Proof Technique for PARO.* Zhen et al. (2018a) introduce FME as both a solution and proof technique for ARO. We apply and extend the latter idea, and use FME to prove worst-case and Pareto optimality of various decision rule structures. We extend and/or generalize known results in ARO, not related to Pareto optimality, and provide more insightful proofs; for example, one that uses FME to establish the results by Bertsimas and Goyal (2012) and Zhen et al. (2018a) on optimality of LDRs under simplex uncertainty sets.

Finally, to better position our contributions vis-à-vis the extant literature, we note that PARO solutions have previously been discussed for a non-linear ARO problem arising in radiation therapy treatment planning (Ten Eikelder et al., 2019), but no general treatment of the topic was included. Herein, we formalize the concept, derive properties, such as existence of PARO solutions, and also discuss constructive approaches towards finding them. With regards to FME, Zhen et al. (2018a) were the first to recognize its applicability to linear ARO problems, owing to its ability to eliminate adaptive variables. They use FME as both a solution and proof technique; for the latter the main obstacle is the exponential increase in number of constraints after variable elimination. In the current paper, we apply and extend the ideas of Zhen et al. (2018a), and use FME as a proof technique. Through the lens of FME we first consider (worst-case) optimality of decision rule structures, and provide more general and more insightful proofs of known results. Subsequently, we investigate Pareto optimality using FME and present numerical results which demonstrate the value of PARO solutions.

Notation and Organization

Boldface characters represent matrices and vectors. All vectors are column vectors and the vector $\boldsymbol{a}_i$ is the i-th row of matrix $\boldsymbol{A}$. The space of all measurable functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ that are bounded on compact sets is denoted by $\mathcal{R}^{n,m}$. The vectors $\boldsymbol{e}_i$, $\mathbf{1}$ and $\mathbf{0}$ are the standard unit basis vector, the vector of all-ones and the vector of all-zeros, respectively. Matrix $\boldsymbol{I}$ is the identity matrix. The relative interior of a set S is denoted by $\text{ri}(S)$; its set of extreme points is denoted by $\text{ext}(S)$.

The manuscript is organized as follows. First, Section 2 introduces the ARO setting and the notion of PARO. Section 3 introduces FME and uses it to establish (worst-case) optimality of decision rule structures. Section 4 investigates the existence of PARO solutions, and Section 5 presents some practical approaches for the construction of PARO solutions. In Section 6 we present the results of our numerical experiments.

2 Pareto Optimality in (Adaptive) Robust Optimization

We first generalize the definition of PRO of Iancu and Trichakis (2014) to non-linear static RO problems. The reason for this is that there turns out to be a relation between Pareto efficiency for non-linear static RO problems and linear ARO problems. Let $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^{n_x}$ denote the decision variables and let $\boldsymbol{z} \in U \subseteq \mathbb{R}^L$ denote the uncertain parameters. Let $f : \mathbb{R}^{n_x} \times \mathbb{R}^L \mapsto \mathbb{R}$ and consider the static RO problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{z}\in U} \; f(\boldsymbol{x},\boldsymbol{z}). \tag{2}$$

Let $\mathcal{X}^{\mathrm{RO}}$ denote the set of robustly optimal (i.e., worst-case optimal) solutions. A robustly optimal solution $\boldsymbol{x}$ is PRO if there does not exist another robustly optimal solution $\bar{\boldsymbol{x}}$ that performs at least as good as $\boldsymbol{x}$ for all scenarios in the uncertainty set, while performing strictly better for at least one scenario. If such a solution $\bar{\boldsymbol{x}}$ does exist, it is said to *dominate* $\boldsymbol{x}$. In practice, solution $\bar{\boldsymbol{x}}$ will always be preferred over $\boldsymbol{x}$. If all uncertainty in the objective is moved to constraints using an epigraph formulation, Pareto robust optimality may also be defined in terms of slack variables (Iancu and Trichakis, 2014, Section 4.1), but we do not use that definition here. We use the following formal definition:

Definition 1 (Pareto Robustly Optimal)**.** A solution $\boldsymbol{x} \in \mathcal{X}^{\mathrm{RO}}$ is PRO to (2) if there does not exist another $\bar{\boldsymbol{x}} \in \mathcal{X}^{\mathrm{RO}}$ such that

$$\begin{aligned} f(\bar{\boldsymbol{x}},\boldsymbol{z}) &\le f(\boldsymbol{x},\boldsymbol{z}), \quad \forall \boldsymbol{z}\in U,\\ f(\bar{\boldsymbol{x}},\bar{\boldsymbol{z}}) &< f(\boldsymbol{x},\bar{\boldsymbol{z}}), \quad \text{for some } \bar{\boldsymbol{z}}\in U. \end{aligned} \qquad \blacksquare$$

We aim to extend the concept of PRO to ARO problems. In particular, we consider the following adaptive linear optimization problem:

$$\min_{\boldsymbol{x},\boldsymbol{y}(\cdot)} \; \max_{\boldsymbol{z}\in U} \; \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{z}), \tag{3a}$$

$$\text{s.t.} \;\; \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \le \boldsymbol{r}(\boldsymbol{z}), \quad \forall \boldsymbol{z}\in U, \tag{3b}$$

where $\boldsymbol{z} \in U \subseteq \mathbb{R}^L$ is an uncertain parameter, with U a compact, convex uncertainty set with nonempty relative interior. Variables $\boldsymbol{x} \in \mathbb{R}^{n_x}$ are the Stage-1 (*here-and-now*) decisions. Usually we will assume $\boldsymbol{x}$ to be continuous variables, but we emphasize that all results in the paper also hold if (part of) $\boldsymbol{x}$ is restricted to be integer-valued. Variables $\boldsymbol{y} \in \mathcal{R}^{L,n_y}$ are also continuous and capture the Stage-2 (*wait-and-see*) decisions, i.e., they are functions of $\boldsymbol{z}$. The matrix $\boldsymbol{B} \in \mathbb{R}^{m\times n_y}$ and vector $\boldsymbol{d} \in \mathbb{R}^{n_y}$ are assumed to be constant (fixed recourse), and $\boldsymbol{A}(\boldsymbol{z})$, $\boldsymbol{r}(\boldsymbol{z})$ and $\boldsymbol{c}(\boldsymbol{z})$ depend affinely on $\boldsymbol{z}$:

$$\boldsymbol{A}(\boldsymbol{z}) := \boldsymbol{A}^0 + \sum_{k=1}^{L} \boldsymbol{A}^k z_k, \quad \boldsymbol{r}(\boldsymbol{z}) := \boldsymbol{r}^0 + \sum_{k=1}^{L} \boldsymbol{r}^k z_k, \quad \boldsymbol{c}(\boldsymbol{z}) := \boldsymbol{c}^0 + \sum_{k=1}^{L} \boldsymbol{c}^k z_k,$$

with $\boldsymbol{A}^0,\dots,\boldsymbol{A}^L \in \mathbb{R}^{m\times n_x}$, $\boldsymbol{r}^0,\dots,\boldsymbol{r}^L \in \mathbb{R}^m$ and $\boldsymbol{c}^0,\dots,\boldsymbol{c}^L \in \mathbb{R}^{n_x}$. Uncertainty in the objective (3a) can be moved to the constraint using an epigraph formulation. Nevertheless, it is explicitly stated in the objective to facilitate a convenient definition of PARO. Let OPT denote the optimal (worst-case) objective value of (3). We continue by stating several assumptions and definitions regarding adaptive robust feasibility and optimality.

Definition 2 (Adaptive Robustly Feasible)**.** A pair $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is Adaptive Robustly Feasible (ARF) to (3) if $\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \leq \boldsymbol{r}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U$. ■

Sometimes it is useful to consider properties of the first- and second-stage decisions separately. Therefore, we also define adaptive robust feasibility for Stage-1 and Stage-2 decisions individually.

Definition 3 (Adaptive Robustly Feasible $\boldsymbol{x}$ and/or $\boldsymbol{y}(\cdot)$)**.**

(i) A Stage-1 decision $\boldsymbol{x}$ is ARF to (3) if there exists a $\boldsymbol{y}(\cdot)$ such that $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is ARF to (3).
(ii) A Stage-2 decision $\boldsymbol{y}(\cdot)$ is ARF to (3) if there exists a $\boldsymbol{x}$ such that $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is ARF to (3). ■

The set of all ARF solutions $\boldsymbol{x}$ is given by

$$\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^{n_x} \mid \exists \boldsymbol{y} \in \mathcal{R}^{L,n_y} : \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \leq \boldsymbol{r}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U\}.$$

We assume set $\mathcal{X}$ is nonempty, i.e., there exists an $\boldsymbol{x}$ that is ARF, and we assume (3) has a finite optimal objective value, i.e., OPT is a finite number. After feasibility, the natural next step is to formally define optimality.

Definition 4 (Adaptive Robustly Optimal)**.** A pair $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is adaptive robustly optimal (ARO)[1] to (3) if it is ARF and $\boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{z}) \leq \text{OPT}, \quad \forall \boldsymbol{z} \in U$. ■

We also define adaptive robust optimality for Stage-1 and Stage-2 decisions individually.

Definition 5 (Adaptive Robustly Optimal $\boldsymbol{x}$ and/or $\boldsymbol{y}(\cdot)$)**.**

(i) A Stage-1 decision $\boldsymbol{x}$ is ARO to (3) if there exists a $\boldsymbol{y}(\cdot)$ such that $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is ARO to (3).
(ii) A Stage-2 decision $\boldsymbol{y}(\cdot)$ is ARO to (3) if there exists a $\boldsymbol{x}$ such that $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is ARO to (3). ■

We are now in position to define Pareto Adaptive Robust Optimality for two-stage ARO problems.

Definition 6 (Pareto Adaptive Robustly Optimal)**.** A pair $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is Pareto Adaptive Robustly Optimal (PARO) to (3) if it is ARO and there does not exist a pair $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\cdot))$ that is ARO and

$$\begin{aligned}
&\boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}(\boldsymbol{z}) \leq \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U, \\
&\boldsymbol{c}(\bar{\boldsymbol{z}})^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}(\bar{\boldsymbol{z}}) < \boldsymbol{c}(\bar{\boldsymbol{z}})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\bar{\boldsymbol{z}}), \quad \text{for some } \bar{\boldsymbol{z}} \in U.
\end{aligned}$$ ■

As before, the definitions can be extended to Stage-1 and Stage-2 decisions individually.

Definition 7 (Pareto Adaptive Robustly Optimal $\boldsymbol{x}$ and/or $\boldsymbol{y}(\cdot)$)**.**

(i) A Stage-1 decision $\boldsymbol{x}$ is PARO to (3) if there exists a $\boldsymbol{y}(\cdot)$ such that $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is PARO to (3).
(ii) A Stage-2 decision $\boldsymbol{y}(\cdot)$ is PARO to (3) if there exists a $\boldsymbol{x}$ such that $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is PARO to (3). ■

Our main interest is in Definition 7(i). The reason for this is that the here-and-now decision $\boldsymbol{x}$ is usually the only one that the decision maker has to commit to. In contrast, instead of using decision rule $\boldsymbol{y}(\cdot)$, one can often resort to re-solving the optimization problem for the second stage once the value of the uncertain parameter has been revealed. This is known as the folding horizon approach, and it is applicable as long as there is time to re-solve between observing $\boldsymbol{z}$ and having to implement $\boldsymbol{y}(\boldsymbol{z})$. There is no such alternative for $\boldsymbol{x}$, however, and different decisions in Stage 1 may lead to different adaptation possibilities in Stage 2.

Lastly, Pareto optimality can also be investigated for Stage-2 decisions if the Stage-1 decision $\boldsymbol{x}$ is fixed.

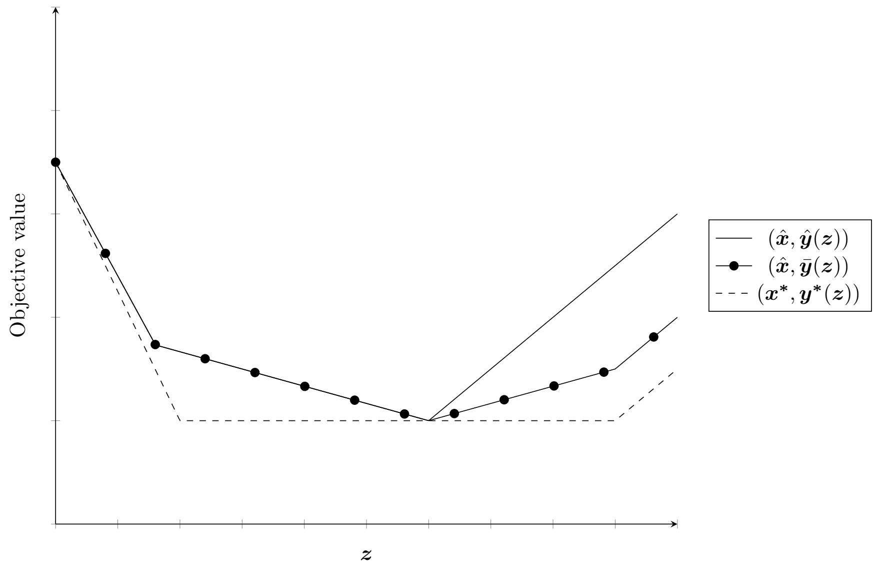


Figure 1: Illustration of PARO concept. Each graph represents the objective value of (3) for a given pair $(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{z}))$ as a function of uncertain parameter $\boldsymbol{z}$. Solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}(\boldsymbol{z}))$ is dominated by $(\hat{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{z}))$. Decision rule $\bar{\boldsymbol{y}}(\boldsymbol{z}))$ may be a PARO extension of $\hat{\boldsymbol{x}}$, decision rule $\hat{\boldsymbol{y}}(\boldsymbol{z}))$ is not. Solution $(\boldsymbol{x}^*, \boldsymbol{y}^*(\boldsymbol{z}))$ dominates both $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}(\boldsymbol{z}))$ and $(\hat{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{z}))$ and may be PARO. Solution $\hat{\boldsymbol{x}}$ may also be a PARO Stage-1 solution.

Definition 8 (Pareto Adaptive Robustly Optimal extension $\boldsymbol{y}(\cdot)$)**.** A Stage-2 decision $\boldsymbol{y}(\cdot)$ is a PARO extension to $\boldsymbol{x}$ for (3) if $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is ARF to (3) and there does not exist another $\bar{\boldsymbol{y}}(\cdot)$ such that $(\boldsymbol{x}, \bar{\boldsymbol{y}}(\cdot))$ is ARF to (3) and

$$\begin{aligned} &\boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}(\boldsymbol{z}) \leq \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U, \\ &\boldsymbol{c}(\bar{\boldsymbol{z}})^\top \boldsymbol{x} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}(\bar{\boldsymbol{z}}) < \boldsymbol{c}(\bar{\boldsymbol{z}})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\bar{\boldsymbol{z}}), \quad \text{for some } \bar{\boldsymbol{z}} \in U. \end{aligned} \qquad \blacksquare$$

Figure 1 illustrates the PARO concept for a single uncertain parameter. Each graph represents the objective value of (3) for a given pair $(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{z}))$ as a function of uncertain parameter $\boldsymbol{z}$. The solution pair $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}(\boldsymbol{z}))$ (solid line) is dominated by $(\hat{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{z}))$ (solid-dotted line), which has the same here-and-now decision $\hat{\boldsymbol{x}}$ but a different decision rule. Thus, according to Definition 8, decision rule $\hat{\boldsymbol{y}}(\cdot)$ cannot be a PARO extension of $\hat{\boldsymbol{x}}$, but decision rule $\bar{\boldsymbol{y}}(\cdot)$ may be. Further, solution $(\boldsymbol{x}^*, \boldsymbol{y}^*(\boldsymbol{z}))$ yields a graph that is below the graphs of the other two solution pairs, so neither of those pairs can be PARO according to Definition 6. However, care must be exercised. It may be the case that there is yet another decision rule $\tilde{\boldsymbol{y}}(\boldsymbol{z})$ so that $(\hat{\boldsymbol{x}}, \tilde{\boldsymbol{y}}(\boldsymbol{z}))$ is not dominated by $(\boldsymbol{x}^*, \boldsymbol{y}^*(\boldsymbol{z}))$. Hence, we cannot conclude that $\hat{\boldsymbol{x}}$ is not PARO. Lastly, $(\boldsymbol{x}^*, \boldsymbol{y}^*(\boldsymbol{z}))$ can be PARO, but that cannot be concluded from the figure either.

We conclude this section by mentioning three ways that the PARO concept can be generalized and relaxed, although we do not consider these any further. First, Bertsimas et al. (2019) define Pareto optimal adaptive solutions for general (non-linear) two-stage ARO problems, which for linear problems is equivalent to our definition of PARO. They subsequently define Pareto optimal affine adaptive solutions, which is equivalent to the definition of PRO after using LDRs, and focus on finding the latter type of solutions. In their numerical studies, Iancu and Trichakis (2014) also consider two-stage problems and find PRO solutions after plugging in LDRs.

Secondly, we can also solely relax the requirement that the solution is ARO. For example, often LDRs do not guarantee an ARO solution but do exhibit good practical performance (Kuhn et al., 2009). Suppose these yield a worst-case objective value p ($>$ OPT). Then we can

[1]To ease exposition, we overload and reuse certain acronyms, such as ARO for "Adaptive Robust Optimization" and "Adaptive Robustly Optimal", as long as they can be readily disambiguated from the context.

define p-PARO solutions as those solutions $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ that yield an objective value of at most p in each scenario, and are not dominated by another solution $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\cdot))$ that yields an objective value of at most p in each scenario.

Thirdly, PARO may also be defined in terms of slack variables, analogous to the extension of PRO to constraint slacks in Iancu and Trichakis (2014, Section 4.1). In that paper, a slack value vector is introduced that quantifies the relative importance of slack in each constraint. This scalarization allows the computation of the total slack value of a solution in any scenario. Subsequently PRO (and also PARO) can be defined on this total slack value instead of the objective value. This may be useful in applications where ARO is mainly used for maintaining feasibility, such as immunizing against uncertain renewable energy source output (Jabr, 2013) and adjusting to disturbances in railway timetabling (Polinder et al., 2019).

3 Optimality of Decision Rule Structures via an FME Lens

In this section, we introduce FME as a proof technique for ARO, and we analyze various decision rule structures. This FME lens enables us to prove that an ARF decision rule with the particular structure exists for *every* ARF $\boldsymbol{x}$, instead of solely proving it is optimal for an ARO $\boldsymbol{x}$. These results are crucial for one of our main results in Section 5. Moreover, the results in this section show that FME not only provides more general results, but also leads to more concise (and perhaps more intuitive) proofs to known results on optimal decision structures.

3.1 Eliminating adaptive variables using Fourier-Motzkin Elimination

FME (Fourier, 1827; Motzkin, 1936) is an algorithm for solving systems of linear inequalities. We refer to Bertsimas and Tsitsiklis (1997) for an introduction to FME in linear optimization. Its usefulness in ARO is due to the fact that it can be used to eliminate adaptive variables, as proposed by Zhen et al. (2018a). The exponential increase in number of constraints is addressed by a redundant constraint identification scheme. Zhen et al. (2018a) also propose to use FME to eliminate only part of the variables and using LDRs for remaining adaptive variables. Next to this, they use FME to prove (worst-case) optimality of PWL decision rules. Furthermore, they consider optimal decision rules for the adaptive variable in the dual problem: they prove (worst-case) optimality of LDRs in case of simplex uncertainty and (two-)piecewise linear decision rules in case of box uncertainty. Zhen and den Hertog (2018) use a combination of FME and ARO techniques to compute the maximum volume inscribed ellipsoid of a polytopic projection. The following example illustrates the use of FME to eliminate an adaptive variable.

Example 2. We use FME to eliminate adaptive variable y from (1) in Example 1. We move the uncertain objective to the constraints using an epigraph variable $t \in \mathbb{R}$, and rewrite the constraints to obtain:

$$\begin{aligned}
\min_{x,t,y(d_1,d_2)} \quad & t, && \text{(4a)}\\
\text{s.t.} \quad & 20 \leq x \leq 40, && \text{(4b)}\\
& y(d_1,d_2) \leq t/\delta - x, \quad \forall (d_1,d_2) \in U, && \text{(4c)}\\
& d_1 - x \leq y(d_1,d_2), \quad \forall (d_1,d_2) \in U, && \text{(4d)}\\
& d_2 - x \leq y(d_1,d_2), \quad \forall (d_1,d_2) \in U, && \text{(4e)}\\
& 20 \leq y(d_1,d_2) \leq 40, \quad \forall (d_1,d_2) \in U. && \text{(4f)}
\end{aligned}$$

For fixed (d_1, d_2), Constraints (4c)-(4f) specify lower and/or upper bounds on y. By combining each pair of lower and upper bounds on y into a new constraint, we find the following problem in terms of (x,t):

$$
\begin{aligned}
\min_{x,t} \quad & t, && \text{(5a)}\\
\text{s.t.} \quad & 20 \le x \le 40, && \text{(5b)}\\
& d_1 \le t/\delta, \quad \forall (d_1,d_2) \in U, && \text{(5c)}\\
& d_2 \le t/\delta, \quad \forall (d_1,d_2) \in U, && \text{(5d)}\\
& 20 \le t/\delta - x, \quad \forall (d_1,d_2) \in U, && \text{(5e)}\\
& d_1 - x \le 40, \quad \forall (d_1,d_2) \in U, && \text{(5f)}\\
& d_2 - x \le 40, \quad \forall (d_1,d_2) \in U, && \text{(5g)}
\end{aligned}
$$

where we have removed the trivial new constraint $20 \le 40$. Any solution (x,t) sets the following bounds on y:

$$\max\{d_1 - x, d_2 - x, 20\} \le y(d_1,d_2) \le \min\{t/\delta - x, 40\}, \quad \forall (d_1,d_2) \in U,$$

and any decision rule satisfying these inequalities is ARO to (1). Thus, two-stage problem (1) has been reduced to static linear RO problem (5). Auxiliary variable t can be eliminated, but this transforms (5) to an RO problem with a PWL objective. ▲

We focus on applying FME as a proof technique. Through the "lens" of FME we first consider (worst-case) optimality of decision rule structures, and subsequenly consider Pareto optimality. In the remainder of the paper, if FME is applied, w.l.o.g. it is applied on the adaptive variables in the order $y_1(\boldsymbol{z}), \ldots, y_{n_y}(\boldsymbol{z})$, i.e., according to their index. We first state some frequently used definitions. If FME is performed on $\mathcal{X}$ until all adaptive variables are eliminated, the feasible region can be written as

$$\mathcal{X}_{\text{FME}} = \{\boldsymbol{x} \in \mathbb{R}^{n_x} \mid \boldsymbol{G}(\boldsymbol{z})\boldsymbol{x} \le \boldsymbol{f}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U\},$$

for some matrix $\boldsymbol{G}(\boldsymbol{z})$ and vector $\boldsymbol{f}(\boldsymbol{z})$ depending affinely on $\boldsymbol{z}$. Zhen et al. (2018a) show that $\mathcal{X} = \mathcal{X}_{\text{FME}}$. For the analysis of particular decision rule structures, it is crucial to keep track of the original constraints during the FME procedure. A frequently used technical result on this is provided in Lemma 10 in Appendix A.1.

3.2 Optimality of decision rule structures

In this section, we consider several special cases of problem (3) for which particular decision rule structures are known to be optimal. We use FME to prove generalizations of these results for linear two-stage ARO problems. In particular, using FME as a proof technique enables us to show that the particular decision rule structure is not only ARO (i.e., worst-case optimal), but is ARF for each ARF Stage-1 decision $\boldsymbol{x}$. These results will later be used in analyzing PARO in Section 5.

We consider the cases where uncertainty appears (i) constraintwise, (ii) in a hybrid structure (part constraintwise, part non-constraintwise), (iii) in a block structure, and we consider (iv) the case with a simplex uncertainty set and the case with only one uncertain parameter.

(i) Constraintwise uncertainty

Ben-Tal et al. (2004) show that for constraintwise uncertainty the objective values of the static and adaptive problem are equal, i.e., there exists an optimal static decision rule. Using FME, a generalization of their result can be easily proved. We first provide an example.

Example 3. Consider the following ARO problem with constraintwise uncertainty:

$$\begin{aligned} \min_{x,\boldsymbol{y}(\cdot)} \quad & x, \\ \text{s.t.} \quad & x - y_2(\boldsymbol{z}) \leq -\frac{1}{2}z_1, \quad \forall z_1 \in [0,1], \\ & -x + y_1(\boldsymbol{z}) + y_2(\boldsymbol{z}) \leq \frac{1}{2}z_2 + \frac{1}{2}z_3 + 2, \quad \forall (z_2, z_3) \in [0,1]^2, \\ & 1 \leq y_1(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U, \\ & \frac{3}{2} \leq y_2(\boldsymbol{z}) \leq 2, \quad \forall \boldsymbol{z} \in U \end{aligned}$$

with $U = [0,1]^3$. Uncertain parameter z_1 occurs only in the first constraint and (z_2, z_3) occur only in the second constraint. Using FME, we first eliminate $y_1(\boldsymbol{z})$ and subsequently eliminate $y_2(\boldsymbol{z})$.

$$\begin{aligned} 1 \leq & y_1(\boldsymbol{z}) \leq -y_2(\boldsymbol{z}) + x + 2 + \frac{1}{2}z_2 + \frac{1}{2}z_3, \\ \max\{\frac{3}{2}, x + \frac{1}{2}z_1\} \leq & y_2(\boldsymbol{z}) \leq \min\{2, x + 1 + \frac{1}{2}z_2 + \frac{1}{2}z_3\}. \end{aligned}$$

One can verify that the (unique) ARO solution is $x^* = \frac{1}{2}$. Additionally, note that the term $\frac{1}{2}z_2 + \frac{1}{2}z_3$ appears in both upper bounds with a positive sign. As this is the only term that depends on (z_2, z_3), it can be replaced by its worst-case value 0. Similarly, the term $-\frac{1}{2}z_1$ appears in the lower bound on $y_2(\boldsymbol{z})$ with a negative sign, and can be replaced by its worst-case value $-\frac{1}{2}$. This yields the following bounds on $y_1(\boldsymbol{z})$ and $y_2(\boldsymbol{z})$:

$$\begin{aligned} 1 \leq & y_1(\boldsymbol{z}) \leq -y_2(\boldsymbol{z}) + \frac{5}{2}, \\ \max\{\frac{3}{2}, 1\} \leq & y_2(\boldsymbol{z}) \leq \min\{2, \frac{3}{2}\}. \end{aligned}$$

For $y_2(\boldsymbol{z})$, the only feasible (and hence ARO) decision rule is $y_2(\boldsymbol{z}) = \frac{3}{2}$. This implies $y_1(\boldsymbol{z}) = 1$, and we find that for both adaptive variables the optimal decision rule is static. ▲

According to Lemma 10, any term such as $\frac{1}{2}z_2 + \frac{1}{2}z_3$ in Example 4 appears in all upper bounds with a positive sign and all lower bounds with a negative sign, or vice versa. Hence, if this is the only term depending on z_2 and z_3, these uncertain parameters can be eliminated by replacing them with their worst-case value. The resulting bounds on adaptive variables are independent of uncertain parameters. Constraintwise uncertainty is formally defined as follows.

Definition 9. ARO problem (3) has constraintwise uncertainty if there is a partition

$$\boldsymbol{z} = (\boldsymbol{z}_{(0)}, \boldsymbol{z}_{(1)}, \ldots, \boldsymbol{z}_{(m)}),$$

such that $\boldsymbol{z}_{(0)}, \ldots, \boldsymbol{z}_{(m)}$ are disjoint, the objective depends only on $\boldsymbol{z}_{(0)}$ and constraint i depends only on $\boldsymbol{z}_{(i)}$, $i = 1, \ldots, m$. Additionally, $U = \{(\boldsymbol{z}_{(0)}, \ldots, \boldsymbol{z}_{(m)}) \mid \boldsymbol{z}_{(i)} \in U^i, i = 0, \ldots, m\}$, with $U^i \subseteq \mathbb{R}^{|\boldsymbol{z}_{(i)}|}$ for all $i = 0, \ldots, m$. ■

Instead of directly providing a formal proof of the result for constraintwise uncertainty, it follows as a corollary from our analysis of hybrid uncertainty, which is considered next.

(ii) Hybrid uncertainty

Hybrid uncertainty is a generalization of constraintwise uncertainty, where part of the uncertain parameters appear constraintwise, and part does not appear constraintwise. This uncertainty structure has previously been considered in Marandi and den Hertog (2018).

In case of hybrid uncertainty, there exist ARO decision rules that do not depend on the constraintwise uncertain parameters. We illustrate this with a toy example.

Example 4. We extend Example 3 to a problem with hybrid uncertainty by introducing a non-constraintwise uncertain parameter $\hat{z}$:

$$\begin{aligned}
\min_{x,\boldsymbol{y}(\cdot)} \quad & x, && \text{(6a)}\\
\text{s.t.} \quad & x - y_2(\boldsymbol{z}) \leq -\hat{z} - \frac{1}{2}z_1, \quad \forall(\hat{z}, z_1) \in [0,1]^2, && \text{(6b)}\\
& -x + y_1(\boldsymbol{z}) + y_2(\boldsymbol{z}) \leq \hat{z} + \frac{1}{2}z_2 + \frac{1}{2}z_3 + 2, \quad \forall(\hat{z}, z_2, z_3) \in [0,1]^3, && \text{(6c)}\\
& 1 \leq y_1(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U, && \text{(6d)}\\
& \frac{3}{2} \leq y_2(\boldsymbol{z}) \leq 2, \quad \forall \boldsymbol{z} \in U, && \text{(6e)}
\end{aligned}$$

with $U = [0,1]^4$. Uncertain parameter $\hat{z}$ occurs in both constraints, z_1 occurs only in the first constraint and (z_2, z_3) occur only in the second constraint. Using FME, we again first eliminate $y_1(\boldsymbol{z})$ and subsequently eliminate $y_2(\boldsymbol{z})$.

$$\begin{aligned}
1 \leq & y_1(\boldsymbol{z}) \leq \hat{z} - y_2(\boldsymbol{z}) + x + 2 + \frac{1}{2}z_2 + \frac{1}{2}z_3,\\
\max\{\frac{3}{2}, x + \hat{z} + \frac{1}{2}z_1\} \leq & y_2(\boldsymbol{z}) \leq \min\{2, x + 1 + \hat{z} + \frac{1}{2}z_2 + \frac{1}{2}z_3\}.
\end{aligned}$$

The unique ARO solution is still $x^* = \frac{1}{2}$. Similar to Example 3, we can replace both occurrences of the term $\frac{1}{2}z_2 + \frac{1}{2}z_3$ by its worst-case value 0, and $-\frac{1}{2}z_1$ can be replaced by its worst-case value $-\frac{1}{2}$. This yields the following bounds on $y_1(\boldsymbol{z})$ and $y_2(\boldsymbol{z})$:

$$\begin{aligned}
1 \leq & y_1(\boldsymbol{z}) \leq \hat{z} - y_2(\boldsymbol{z}) + \frac{5}{2},\\
\max\{\frac{3}{2}, 1 + \hat{z}\} \leq & y_2(\boldsymbol{z}) \leq \min\{2, \frac{3}{2} + \hat{z}\}.
\end{aligned}$$

For $y_2(\boldsymbol{z})$, the only feasible (and hence ARO) LDR is $y_2(\hat{z}) = \frac{3}{2} + \frac{1}{2}\hat{z}$. This implies $1 \leq y_1(\boldsymbol{z}) \leq 1 + \frac{1}{2}\hat{z}$, and any decision rule that satisfies these bounds is ARO. Note that both decision rules do not depend on the constraintwise uncertain parameters. One can also pick a PWL decision rule for $y_2(\boldsymbol{z})$, such as its lower or upper bound. Also in this case the decision rules for y_1 and y_2 do not depend on z_1, z_2 or z_3. ▲

Hybrid uncertainty is defined as follows.

Definition 10. ARO problem (3) has hybrid uncertainty if there is a partition

$$\boldsymbol{z} = (\hat{\boldsymbol{z}}, \boldsymbol{z}_{(0)}, \boldsymbol{z}_{(1)}, \ldots, \boldsymbol{z}_{(m)}),$$

such that $\hat{\boldsymbol{z}}, \boldsymbol{z}_{(0)}, \ldots, \boldsymbol{z}_{(m)}$ are disjoint, the objective depends only on $\hat{\boldsymbol{z}}$ and $\boldsymbol{z}_{(0)}$ and constraint i depends only on $\hat{\boldsymbol{z}}$ and $\boldsymbol{z}_{(i)}$, $i = 1, \ldots, m$. Additionally, $U = \{(\hat{\boldsymbol{z}}, \boldsymbol{z}_{(0)}, \ldots, \boldsymbol{z}_{(m)}) \mid \hat{\boldsymbol{z}} \in \hat{U}, \boldsymbol{z}_{(i)} \in U^i, i = 0, \ldots, m\}$, with $\hat{U} \subseteq \mathbb{R}^{|\hat{\boldsymbol{z}}|}$ and $U^i \subseteq \mathbb{R}^{|\boldsymbol{z}_{(i)}|}$ for all $i = 0, \ldots, m$. ■

To formally prove our claim that there exist ARO decision rules that do not depend on the constraintwise uncertain parameters, we first need a result on feasibility.

Lemma 1. *Let P_{hybrid} denote an ARO problem of form* (3) *with hybrid uncertainty and let $\boldsymbol{x}$ be ARF to P_{hybrid}. Then, there exists a decision rule $\boldsymbol{y}(\cdot)$ that depends only on $\hat{\boldsymbol{z}}$ such that $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is ARF to P_{hybrid}.*

Proof. See Appendix A.2. □

The following result is an immediate consequence of Lemma 1 for ARO decisions.

Corollary 1. *Let P_{hybrid} denote an ARO problem of form* (3) *with hybrid uncertainty. For each $\boldsymbol{x}$ that is ARO to P_{hybrid} there exists a decision rule $\boldsymbol{y}(\cdot)$ depending only on $\hat{\boldsymbol{z}}$ such that the pair $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is ARO to P_{hybrid}.*

Proof. See Appendix A.3. □

In case of pure constraintwise uncertainty ($U^0 = \emptyset$) Lemma 1 shows that for each ARF $\boldsymbol{x}$ there exists a static $\boldsymbol{y}$ such that $(\boldsymbol{x}, \boldsymbol{y})$ is ARF. Additionally, Corollary 1 shows that for each ARO $\boldsymbol{x}$ there exists a static $\boldsymbol{y}$ such that $(\boldsymbol{x}, \boldsymbol{y})$ is ARO.

Marandi and den Hertog (2018) prove a similar result to Corollary 1 for non-linear problems. More precisely, they prove that for problems with hybrid uncertainty there exists an optimal decision rule that is a function of only the non-constraintwise uncertain parameters if the problem is convex in the decision variables, concave in uncertain parameters, has a convex compact uncertainty set and a convex compact feasible region for the adaptive variables.

(iii) Block uncertainty

Suppose we can split the constraints into blocks, where each block has its own uncertain parameters and adaptive variables, and the uncertainty set is a Cartesian product of the block-wise uncertainty sets, then there exists an optimal decision rule for each adaptive variable that depends only on the uncertain parameters in its own block. We first provide an example to develop some intuition for block uncertainty.

Example 5. Consider again Example 4. Add the following constraints to (6):

$$y_3(\boldsymbol{z}) + x \leq -\frac{1}{2}z_4 + \frac{3}{2}, \quad \forall z_4 \in [0,1], \tag{7a}$$

$$y_3(\boldsymbol{z}) + 2x \geq \frac{1}{2}z_5 + 1, \quad \forall z_5 \in [0,1]. \tag{7b}$$

Then the first block consists of constraints (6b)-(6e), adaptive variables $y_1(\boldsymbol{z}), y_2(\boldsymbol{z})$ and uncertain parameters $z_0, \dots, z_3$. The second block consists of constraints (7), adaptive variable $y_3(\boldsymbol{z})$ and uncertain parameters z_4 and z_5. One can verify that the unique ARO solution remains $x^* = \frac{1}{2}$. The following bounds on $y_3(\boldsymbol{z})$ are obtained:

$$\frac{1}{2}z_5 \leq y_3(\boldsymbol{z}) \leq 1 - \frac{1}{2}z_4.$$

One feasible (and hence ARO) decision rule is $y_3(z_4, z_5) = \frac{1}{2}(1 + z_5 - z_4)$. The decision rules for y_1 and y_2 remain unchanged. It follows that for each adaptive variable the optimal decision rule is a function of only the uncertain parameters it its own block. ▲

The formal definition of block uncertainty is as follows. Recall that constraints are indexed $1, \dots, m$. Let index 0 refer to the objective.

Definition 11. ARO problem (3) has block uncertainty if there exist partitions $\boldsymbol{z} = (\boldsymbol{z}_{(1)}, \dots, \boldsymbol{z}_{(V)})$, $\boldsymbol{y}(\cdot) = (\boldsymbol{y}_{(1)}(\cdot), \dots, \boldsymbol{y}_{(V)}(\cdot))$ and $\{0, \dots, m\} = \{K_{(1)}, \dots, K_{(V)}\}$ such that

- $U = \{(\boldsymbol{z}_{(1)}, \dots, \boldsymbol{z}_{(V)}) \mid \boldsymbol{z}_{(v)} \in U^v, v = 1, \dots, V\}$, with $U^v \subseteq \mathbb{R}^{|\boldsymbol{z}_{(v)}|}$ for all blocks $v = 1, \dots, V$.
- A constraint or objective with index in set $K_{(v)}$ is independent of uncertain parameters $\boldsymbol{z}_{(w)}$ and adaptive variables $\boldsymbol{y}_{(w)}$ if block $w \neq v$. ■

In order to prove the claim that there exists an optimal decision rule for each adaptive variable that depends only on the uncertain parameters in its own block, we again first consider feasibility.

Lemma 2. *Let P_{block} denote an ARO problem of form (3) with block uncertainty and let $\boldsymbol{x}$ be ARF to P_{block}. Then there exists a decision rule $\boldsymbol{y}(\cdot)$ with $\boldsymbol{y}_{(v)}(\cdot)$ depending only on $\boldsymbol{z}_{(v)}$, for all $v = 1, \dots, V$, such that $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is ARF to P_{block}.*

Proof. See Appendix A.4. □

Corollary 2. *Let P_{block} denote an ARO problem of form* (3) *with block uncertainty. For each $\boldsymbol{x}$ that is ARO to P_{block} there exists a decision rule $\boldsymbol{y}(\cdot)$ with $\boldsymbol{y}_{(v)}(\cdot)$ depending only on $\boldsymbol{z}_{(v)}$, for all $v = 1, \dots, V$, such that the pair $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is ARO to P_{block}.*

Proof. Follows from Lemma 2 analogous to the proof of Corollary 1. □

(iv) Simplex uncertainty or one uncertain parameter

Bertsimas and Goyal (2012) prove optimality of LDRs for right-hand side uncertainty and a simplex uncertainty set. Zhen et al. (2018a) generalize this to both left- and right-hand side uncertainty, their proof uses FME on the dual problem. We use FME on the primal problem, which leads to a more intuitive proof; the following example illustrates the main idea. We note that the case with one uncertain parameter is a special case of simplex uncertainty, so the results of this section also hold for that case.

Example 6. Consider the problem

$$\begin{aligned}
\min \quad & x, \\
\text{s.t.} \quad & x - y_2 \leq -z_1 - \frac{1}{2}z_2 - \frac{1}{2}, \quad \forall \boldsymbol{z} \in U, \\
& -x + y_1 + y_2 \leq z_1 + z_3 + 2, \quad \forall \boldsymbol{z} \in U, \\
& 0 \leq y_1(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U, \\
& \frac{3}{2} \leq y_2(\boldsymbol{z}) \leq 2, \quad \forall \boldsymbol{z} \in U,
\end{aligned}$$

with standard simplex uncertainty set $U = \{(z_1, z_2, z_3) : z_1 + z_2 + z_3 \leq 1, z_1, z_2, z_3 \geq 0\}$. Similar to Example 4, we first eliminate y_1 and then y_2. This yields the following bounds on the adaptive variables:

$$0 \leq y_1 \leq z_1 + z_3 + 2 + x - y_2, \tag{8a}$$

$$\max\{\frac{3}{2}, \frac{1}{2} + x + z_1 + \frac{1}{2}z_2\} \leq y_2 \leq \min\{2, z_1 + z_3 + 1 + x\}, \tag{8b}$$

and these bounds have to be satisfied for each point in $\text{ext}(U)$. One can verify that $x^* = \frac{1}{2}$ is an ARO solution. Plugging this in (8), we get the following bounds for each extreme point:

$$\begin{aligned}
(0,0,0): \quad & 0 \leq y_1 \leq \frac{5}{2} - y_2, \quad \frac{3}{2} \leq y_2 \leq \frac{3}{2}, \\
(1,0,0): \quad & 0 \leq y_1 \leq \frac{7}{2} - y_2, \quad 2 \leq y_2 \leq 2, \\
(0,1,0): \quad & 0 \leq y_1 \leq \frac{5}{2} - y_2, \quad \frac{3}{2} \leq y_2 \leq \frac{3}{2}, \\
(0,0,1): \quad & 0 \leq y_1 \leq \frac{7}{2} - y_2, \quad \frac{3}{2} \leq y_2 \leq 2.
\end{aligned} \tag{9}$$

Because U is a simplex, the four extreme points are affinely independent. Therefore, there is a unique LDR such that the upper bound on $y_2(\cdot)$ holds with equality for each extreme point. This is also the case for the lower bound, and any convex combination of both decision rules also satisfies the bounds for y_2 in (9). The LDR corresponding with the upper bounds is $y_2(z_1, z_3) = \frac{1}{2}(3 + z_1 + z_3)$, and plugging this in the bounds on y_1 yields a similar system as (9) for y_1. This guarantees existence of an LDR for y_1; for the upper bound we find $y_1(z_1, z_3) = \frac{1}{2}(2 + z_1 + z_3)$. Note that this does not generalize to uncertainty sets described by more than $L + 1$ extreme points. ▲

Similar to the cases for hybrid and block uncertainty, we first prove feasibility for each ARF $\boldsymbol{x}$, and subsequently prove optimality.

Lemma 3. *Let $P_{simplex}$ denote an ARO problem of form* (3) *with a simplex uncertainty set, i.e., $U = Conv(\boldsymbol{z}^1, \dots, \boldsymbol{z}^{L+1})$, with $\boldsymbol{z}^j \in \mathbb{R}^L$ such that $\boldsymbol{z}^1, \dots, \boldsymbol{z}^{L+1}$ are affinely independent. Let $\boldsymbol{x}$ be ARF to $P_{simplex}$. Then there exists an LDR $\boldsymbol{y}(\cdot)$ such that $(\boldsymbol{x}, \boldsymbol{y})$ is ARF to $P_{simplex}$.*

Proof. See Appendix A.5. ☐

Similar to Corollary 1, we have the following result for ARO decisions.

Corollary 3. *Let $P_{simplex}$ denote an ARO problem of form* (3) *with a simplex uncertainty set, i.e., $U = Conv(\boldsymbol{z}^1, \ldots, \boldsymbol{z}^{L+1})$, with $\boldsymbol{z}^j \in \mathbb{R}^L$ such that $\boldsymbol{z}^1, \ldots, \boldsymbol{z}^{L+1}$ are affinely independent. For each $\boldsymbol{x}$ that is ARO to $P_{simplex}$ there exists an LDR $\boldsymbol{y}(\cdot)$ such that the pair $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is ARO to $P_{simplex}$.*

Proof. Follows from Lemma 3 analogous to the proof of Corollary 1. ☐

Because the case with one uncertain parameter is a special case of simplex uncertainty, the results of Lemma 3 and Corollary 3 also hold for that case.

The results on PARO in the next sections make use of the fact that an ARF decision rule with the particular structure exist for *every* ARF $\boldsymbol{x}$, i.e., Lemmas 1 to 3.

4 Properties of PARO Solutions

In this section, we prove existence of PARO solutions for two-stage ARO problems of form (3). First, we use FME to prove that a PARO Stage-1 (here-and-now) solution is equivalent to a PRO solution of a PWL convex static RO problem, and use that to prove the existence of a PARO Stage-1 solution. Subsequently, we prove that there exists a PWL decision rule that is PARO to (3).

4.1 Existence of a PARO Stage-1 solution

We prove existence of PARO Stage-1 solutions in two steps. First, we prove that a PARO solution to (3) is equivalent to a PRO solution to a static RO problem with a convex PWL objective. Subsequently, we prove that PRO solutions to such problems always exist.

Lemma 4. *A solution $\boldsymbol{x}^*$ is PARO to* (3) *if and only if it is PRO to*

$$\min_{\boldsymbol{x} \in \mathcal{X}_{FME}} \quad \max_{\boldsymbol{z} \in U} \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \max_{(S,T) \in M} \{h_{S,T}(\boldsymbol{x}, \boldsymbol{z})\}, \tag{10}$$

where each element (S,T) of set M is a pair of sets of original constraints of (3) *and each function $h_{S,T}(\boldsymbol{x}, \boldsymbol{z})$ is bilinear in $\boldsymbol{x}$ and $\boldsymbol{z}$.*

Proof. See Appendix A.6. ☐

Thus, existence of a PARO solution to (3) is now reduced to existence of a PRO solution to a static RO problem with a convex PWL objective in both $\boldsymbol{x}$ and $\boldsymbol{z}$. For problems without adaptive variables in the objective the following result immediately follows.

Corollary 4. *If $\boldsymbol{d} = \boldsymbol{0}$, a solution $\boldsymbol{x}^*$ is PARO to* (3) *if and only if it is PRO to*

$$\min_{\boldsymbol{x} \in \mathcal{X}_{FME}} \max_{\boldsymbol{z} \in U} \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x}.$$

Proof. This directly follows from plugging in $\boldsymbol{d} = \boldsymbol{0}$ in the proof of Lemma 4. ☐

We can now prove one of our main results: existence of a PARO $\boldsymbol{x}$ for any ARO problem of form (3) with compact feasible region. Our proof uses Lemma 4 and essentially proves existence of a PRO solution to (10).

Theorem 1. *If $\mathcal{X}$ is compact, there exists a PARO $\boldsymbol{x}$ to* (3)*.*

Proof. See Appendix A.7. ☐

Note that the theorem also holds if $\mathcal{X}$ restricts (some elements of) $\boldsymbol{x}$ to be integer-valued.

4.2 Existence of a PARO piecewise linear decision rule

Now that existence of a PARO $\boldsymbol{x}$ is established, we investigate the structure of decision rule $\boldsymbol{y}(\cdot)$. We illustrate via an example that for any ARF $\boldsymbol{x}$ there exists a PWL PARO extension $\boldsymbol{y}(\cdot)$.

Example 7. Consider the following ARO problem, a slight adaptation of Example 4:

$$\min_{x,\boldsymbol{y}(\cdot)} \max_{\boldsymbol{z}\in[0,1]^4} x - y_1(\boldsymbol{z}) + y_2(\boldsymbol{z}), \tag{11a}$$

$$\text{s.t. } x - y_2(\boldsymbol{z}) \leq -z_0 - \frac{1}{2}z_1, \quad \forall (z_0, z_1) \in [0,1]^2, \tag{11b}$$

$$-x + y_1(\boldsymbol{z}) + y_2(\boldsymbol{z}) \leq z_0 + \frac{1}{2}z_2 + \frac{1}{2}z_3 + 2, \quad \forall (z_0, z_2, z_3) \in [0,1]^3, \tag{11c}$$

$$1 \leq y_1(\boldsymbol{z}) \leq 2, \quad \forall \boldsymbol{z} \in U, \tag{11d}$$

$$\frac{3}{2} \leq y_2(\boldsymbol{z}) \leq 2, \quad \forall \boldsymbol{z} \in U. \tag{11e}$$

We eliminate $y_1(\boldsymbol{z})$ and $y_2(\boldsymbol{z})$ in constraints (11b)-(11e) analogous to Example 4, and find the ARF solution $x^* = \frac{1}{2}$ and the following bounds on $y_1(\boldsymbol{z})$ and $y_2(\boldsymbol{z})$:

$$1 \leq y_1(\boldsymbol{z}) \leq \min\{2, z_0 - y_2(\boldsymbol{z}) + \frac{5}{2}\}, \tag{12a}$$

$$\max\{\frac{3}{2}, 1 + z_0\} \leq y_2(\boldsymbol{z}) \leq \min\{2, \frac{3}{2} + z_0\}. \tag{12b}$$

Variables $y_1(\boldsymbol{z})$ and $y_2(\boldsymbol{z})$ have not been eliminated in the objective. Therefore, any decision rule satisfying (12) is ARF to (11) but need not be ARO or PARO.

Variable $y_1(\boldsymbol{z})$ does not appear in the bounds of $y_2(\boldsymbol{z})$, so we can consider its individual contribution to the objective value. The objective coefficient of $y_1(\boldsymbol{z})$ is negative, so for any $\boldsymbol{z}$ (including the worst-case) the best possible contribution of $y_1(\boldsymbol{z})$ to the objective value is achieved if we set $y_1(\boldsymbol{z})$ equal to its upper bound. Therefore, for the given x^*, we have the following PWL PARO extension as a function of $y_2(\boldsymbol{z})$:

$$y_1^*(\boldsymbol{z}) = \min\{2, z_0 - y_2(\boldsymbol{z}) + \frac{5}{2}\}.$$

Now that $y_1(\boldsymbol{z})$ is eliminated in the objective value, it remains to find the optimal decision rule for $y_2(\boldsymbol{z})$. Variable $y_2(\boldsymbol{z})$ now appears directly in the objective (11a) and through its occurence in the decision rule $y_1^*(\boldsymbol{z})$. For fixed $\boldsymbol{z}$, the optimal $y_2(\boldsymbol{z})$ is determined by solving

$$\min_{y_2(\boldsymbol{z})} - \min\{2, z_0 - y_2(\boldsymbol{z}) + \frac{5}{2}\} + y_2(\boldsymbol{z}),$$
$$\text{s.t. } \max\{\frac{3}{2}, 1 + z_0\} \leq y_2(\boldsymbol{z}) \leq \min\{2, \frac{3}{2} + z_0\}.$$

One can easily see that the objective is increasing in $y_2(\boldsymbol{z})$, so for any $\boldsymbol{z}$ the best possible contribution of $y_2(\boldsymbol{z})$ to the objective value is achieved if we set $y_2(\boldsymbol{z})$ equal to its lower bound. Thus, for the given x^*, we have the following PWL PARO extension:

$$y_2^*(\boldsymbol{z}) = \max\{\frac{3}{2}, 1 + z_0\}.$$

Note that plugging in a PWL argument in a PWL function retains the piecewise linear structure. Therefore, we also obtain the following PWL PARO extension for $y_1^*(\boldsymbol{z})$:

$$y_1^*(\boldsymbol{z}) = \min\{2, z_0 - \max\{\frac{3}{2}, 1 + z_0\} + \frac{5}{2}\}.$$

Note that we did not move adaptive variables in the objective to the constraints using an epigraph variable, as was done in Example 2. Using an epigraph variable for the objective ensures that each decision rule satisfying the bounds is worst-case optimal, but prevents from comparing performance in other scenarios. Naturally, computationally it has the major advantage that it remains a linear program. ▲

Bemporad et al. (2003) show worst-case optimality of PWL decision rules for right-hand polyhedral uncertainty, i.e., ARO PWL decision rules in our terminology. Zhen et al. (2018a, Theorem 3) generalize this to problems of form (3) with particular assumptions on the uncertainty set. These decision rules are general PWL in $\boldsymbol{z}$ for all variables y_j, $j \neq l$, where y_l is the last eliminated variable in the FME procedure. The decision rule is convex or concave PWL in y_l. These results solely consider the performance of PWL decision rules in the worst-case. Example 7 illustrates that for any ARF $\boldsymbol{x}$ there exists a PWL PARO extension $\boldsymbol{y}(\cdot)$. The lemma below formalizes this claim.

Lemma 5. *For any $\boldsymbol{x}$ that is ARF to* (3) *there exists a PARO extension $\boldsymbol{y}(\boldsymbol{z})$ that is PWL in $\boldsymbol{z}$.*

We present two proofs to Lemma 5; one via FME using the idea of Example 7, and one via basic solutions in linear optimization.

Proof of Lemma 5 via FME.. See Appendix A.8. ☐

Proof of Lemma 5 via linear optimization. See Appendix A.9. ☐

In both proofs the constructed decision rule is in fact optimal for *all* scenarios in the uncertainty set. As long as $\boldsymbol{x}$ is fixed, this is necessary for PARO solutions. The following theorem establishes the existence of PARO PWL decision rules.

Theorem 2. *If $\mathcal{X}$ is compact, there exists a PARO $\boldsymbol{y}(\cdot)$ for* (3) *such that $\boldsymbol{y}(\boldsymbol{z})$ is PWL in $\boldsymbol{z}$.*

Proof. According to Theorem 1 there exists a PARO $\boldsymbol{x}$, and according to Lemma 5 there exists a PARO extension $\boldsymbol{y}(\cdot)$ for this $\boldsymbol{x}$ that is PWL in $\boldsymbol{z}$. It immediately follows that $\boldsymbol{y}(\cdot)$ is PARO. ☐

5 Constructing PARO Solutions

Adaptive robust optimization problems of form (3) are in general NP-hard (Guslitser, 2002), and finding ARO solutions is still the focus of ongoing research (Yanıkoğlu et al., 2019). Thus, finding a method that, given an ARO solution to (3), can produce a PARO solution is not an easy task either. The methods used in the existence proofs of Section 4 are not computationally tractable, i.e., they provide little guidance for finding PARO solutions in practice. In this section we present several practical methods for finding and approximating PARO solutions for particular problems.

First, we consider the problems with known ARO decision rules of Section 3, and show how to obtain PARO solutions in case Stage-2 variables do not appear in the objective. Subsequently, we show how for fixed $\boldsymbol{x}$ we can check whether $\boldsymbol{y}(\cdot)$ is a PARO extension. After that, we consider an application of the finite subset approach of Hadjiyiannis et al. (2011). Lastly, we consider two practical approaches for finding (approximate) PARO solutions if a convex hull description of the uncertainty set is available.

5.1 Known worst-case optimal decision rules

In Section 3, we have shown that for particular ARO problems there exist decision rule structures such that for any ARF Stage-1 decision there exists an ARF decision rule with that structure. For example, for ARO problems with hybrid uncertainty, for any ARF Stage-1 decision there exists an ARF decision that depends only on the non-constraintwise uncertain parameter. It turns out that, in case there are no adaptive variables in the objective, PRO solutions to the static problem obtained after plugging in that decision rule structure are PARO solutions to the original ARO problem. To formalize this, let $\boldsymbol{y}(\boldsymbol{z}) = f_{\boldsymbol{w}}(\boldsymbol{z})$ be a decision rule with known form f (e.g., linear or quadratic) and finite number of parameters $\boldsymbol{w} \in \mathbb{R}^p$, such that $f_{\boldsymbol{w}}(\boldsymbol{z}) \in \mathcal{R}^{L,n_y}$ for any $\boldsymbol{w}$.

Theorem 3. *Let P denote an ARO problem of form* (3) *with $\boldsymbol{d} = \boldsymbol{0}$ and where for any ARF $\boldsymbol{x}$ there exists an ARF decision rule of form $\boldsymbol{y}^*(\boldsymbol{z}) = f_{\boldsymbol{w}}(\boldsymbol{z})$ for some $\boldsymbol{w}$. Then any $\boldsymbol{x}^*$ that is PRO to the static robust optimization problem obtained after plugging in decision rule structure $f_{\boldsymbol{w}}(\boldsymbol{z})$ is PARO to P.*

Proof. See Appendix A.10. ☐

Due to Lemmas 1 to 3, the following result immediately follows for hybrid, block and simplex uncertainty.

Corollary 5.

(i) Let P_{hybrid} denote an ARO problem of form (3) *with $\boldsymbol{d} = \boldsymbol{0}$ and hybrid uncertainty. Let Q denote the static robust optimization problem obtained from P_{hybrid} by plugging in a decision rule structure that depends only on the non-constraintwise parameter. Any $\boldsymbol{x}^*$ that is PRO to Q is PARO to P_{hybrid}.*
(ii) Let P_{block} denote an ARO problem of form (3) *with $\boldsymbol{d} = \boldsymbol{0}$ and block uncertainty. Let Q denote the static robust optimization problem obtained from P_{block} by plugging in a decision rule structure where adaptive variable $\boldsymbol{y}^*_{(v)}(\cdot)$ depend only on $\boldsymbol{z}_{(v)}$ for all $v = 1, \dots, V$. Then any $\boldsymbol{x}^*$ that is PRO to Q is PARO to P_{block}.*
(iii) Let $P_{simplex}$ denote an ARO problem of form (3) *with $\boldsymbol{d} = \boldsymbol{0}$ and a simplex uncertainty set, i.e., $U = Conv(\boldsymbol{z}^1, \dots, \boldsymbol{z}^{L+1})$, with $\boldsymbol{z}^j \in \mathbb{R}^L$ such that $\boldsymbol{z}^1, \dots, \boldsymbol{z}^{L+1}$ are affinely independent. Let Q denote the static robust optimization problem obtained from $P_{simplex}$ by plugging in an LDR structure. Then any $\boldsymbol{x}^*$ that is PRO to Q is PARO to $P_{simplex}$.*

Proof. See Appendix A.11. ☐

Similar to Section 3, the case with constraintwise uncertainty is again a special case of Corollary 5(i). The case with one uncertain parameter is again a special case of Corollary 5(iii). Note that, unlike for worst-case optimization, it is necessary that $\boldsymbol{d} = \boldsymbol{0}$, because our definition of PRO involves the term $\boldsymbol{d}$. If $\boldsymbol{d} \neq \boldsymbol{0}$, the results above does not hold. This is also illustrated in Example 1 in Section 1.

The results of Corollary 5 can be combined. For example, for problems with both simplex uncertainty and hybrid uncertainty, Corollary 5(i) and Corollary 5(iii) together imply that one needs to consider only decision rules that are affine in the non-constraintwise parameter, if there are no adaptive variables in the objective. Simplex uncertainty sets arise in a variety of applications and can be used to approximate other uncertainty sets (Ben-Tal et al., 2020).

5.2 Check whether a decision rule is a PARO extension

If the Stage-1 decision $\boldsymbol{x}$ is fixed, one can verify whether the decision rule $\boldsymbol{y}$ is a PARO extension (Definition 8) as follows.

Lemma 6. *Let $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}(\cdot))$ be an ARO solution to* (3)*. Consider the problem*

$$\max_{\boldsymbol{z},\boldsymbol{y}} \quad \boldsymbol{d}^\top(\tilde{\boldsymbol{y}}(\boldsymbol{z}) - \boldsymbol{y}), \tag{13a}$$

$$s.t. \quad \boldsymbol{A}(\boldsymbol{z})\tilde{\boldsymbol{x}} + \boldsymbol{B}\boldsymbol{y} \leq \boldsymbol{r}(\boldsymbol{z}), \tag{13b}$$

$$\boldsymbol{z} \in U. \tag{13c}$$

If the optimal objective value is zero, $\tilde{\boldsymbol{y}}(\cdot)$ is a PARO extension of $\tilde{\boldsymbol{x}}$. If the objective value is positive, then $\tilde{\boldsymbol{y}}(\cdot)$ is not a PARO extension of $\tilde{\boldsymbol{x}}$ and the suboptimality of $\tilde{\boldsymbol{y}}(\cdot)$ is bounded by the optimal objective value.

Proof. See Appendix A.12. ☐

If the optimal value is positive and $(\boldsymbol{z}^*, \boldsymbol{y}^*)$ denotes an optimal solution to (13), then $\boldsymbol{z}^*$ is a scenario where the suboptimality bound is attained, and $\boldsymbol{y}^*$ is an optimal decision for this scenario. Also, note that if the optimal value of (13) equals zero, the pair $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}(\cdot))$ need not be PARO; there may be a different pair $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}(\cdot))$ that dominates the current pair.

5.3 Unique ARO solution on finite subset of scenarios is PARO

The finite subset approach of Hadjiyiannis et al. (2011) can be used in a PARO setting as well. If the lower bound problem has a unique solution and this solution is feasible to the original problem, it is a PARO solution to the original problem. This is formalized in Lemma 7.

Lemma 7. *Let $S = \{\boldsymbol{z}^1, \ldots, \boldsymbol{z}^N\}$ denote a finite set of scenarios, $S \subseteq U$. Let $\boldsymbol{x}$ be the unique ARO here-and-now decision for which there exist $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^N$ such that $(\boldsymbol{x}, \boldsymbol{y}^1, \ldots, \boldsymbol{y}^N)$ are an optimal solution to*

$$\min_{\boldsymbol{x}, \boldsymbol{y}^1, \ldots, \boldsymbol{y}^N} \quad \max_{i=1,\ldots,N} \{\boldsymbol{c}(\boldsymbol{z}^i)^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}^i\}, \tag{14a}$$

$$\text{s.t.} \quad \boldsymbol{A}(\boldsymbol{z}^i)\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^i \leq \boldsymbol{r}(\boldsymbol{z}^i), \ \forall i = 1, \ldots, N. \tag{14b}$$

Then $\boldsymbol{x}$ is PARO to (3)*.*

Proof. Let $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\cdot))$ be ARO to (3) with $\bar{\boldsymbol{x}}$ unequal to $\boldsymbol{x}$. Then the solution $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\boldsymbol{z}^1), \ldots, \bar{\boldsymbol{y}}(\boldsymbol{z}^N))$ is feasible to (14). Because $\boldsymbol{x}$ is the unique here-and-now ARO decision that can be extended to an optimal solution of (14), it holds that

$$\boldsymbol{c}(\boldsymbol{z}^i)^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}^i < \boldsymbol{c}(\boldsymbol{z}^i)^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}(\boldsymbol{z}^i) \text{ for some } \boldsymbol{z}^i \in S.$$

That is, for each $\bar{\boldsymbol{x}}$ that is ARO to (3) and unequal to $\boldsymbol{x}$ there is at least one scenario $\boldsymbol{z}^i$ in U for which $\boldsymbol{x}$ outperforms $\bar{\boldsymbol{x}}$. This implies that $\boldsymbol{x}$ is PARO to (3). ☐

It should be noted that requiring $\boldsymbol{x}$ to be both ARO to (3) and a unique optimal solution to (14) is quite restrictive.

5.4 Convex hull description of scenario set

Next, consider the case where the uncertainty set is given by the convex hull of a finite set of points, i.e., $U = \text{Conv}(\boldsymbol{z}^1, \ldots, \boldsymbol{z}^N)$. Then (3) is equivalent to

$$\min_{\boldsymbol{x}, \boldsymbol{y}^1, \ldots, \boldsymbol{y}^N} \quad \max_{i=1,\ldots,N} \boldsymbol{c}(\boldsymbol{z}^i)^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}^i, \tag{15a}$$

$$\text{s.t.} \quad \boldsymbol{A}(\boldsymbol{z}^i)\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^i \leq \boldsymbol{r}(\boldsymbol{z}^i), \ \forall i = 1, \ldots, N. \tag{15b}$$

Let $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}^{\boldsymbol{1}}, \ldots, \tilde{\boldsymbol{y}}^{\boldsymbol{N}})$ denote the optimal solution. Note that $\tilde{\boldsymbol{x}}$ is ARO to (3). Analogous to Iancu and Trichakis (2014), we can perform an additional step by optimizing the set of ARO solutions over a scenario in the relative interior of the convex hull of our finite set of scenarios. If the objective does not contain adaptive variables, this yields a PARO here-and-now solution to (3).

Lemma 8. *Let $\boldsymbol{d} = \boldsymbol{0}$. Let $U = Conv(\boldsymbol{z}^1, \ldots, \boldsymbol{z}^N)$, $\bar{\boldsymbol{z}} \in ri(U)$ and let $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}^1, \ldots, \bar{\boldsymbol{y}}^N)$ denote an optimal solution to*

$$\min_{\boldsymbol{x}, \boldsymbol{y}^1, \ldots, \boldsymbol{y}^N} \quad \boldsymbol{c}(\bar{\boldsymbol{z}})^\top \boldsymbol{x}, \tag{16a}$$

$$\text{s.t.} \quad \boldsymbol{A}(\boldsymbol{z}^i)\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^i \leq \boldsymbol{r}(\boldsymbol{z}^i), \ \forall i = 1, \ldots, N, \tag{16b}$$

$$\boldsymbol{c}(\boldsymbol{z}^i)^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}^i \leq OPT, \ \forall i = 1, \ldots, N, \tag{16c}$$

where OPT denotes the optimal objective value of (15)*. Then $\bar{\boldsymbol{x}}$ is PARO to* (3)*.*

Proof. See Appendix A.13. □

For the general case with $\boldsymbol{d} \neq \boldsymbol{0}$, we restrict ourselves to problems with right-hand side uncertainty. Let $\hat{\boldsymbol{x}}$ denote an ARO (worst-case optimal) solution. Let V denote a set where each element is a pair of a scenario in U and a required objective value for that scenario. We initialize $V = \{(\boldsymbol{z}^1, \text{OPT}), \ldots, (\boldsymbol{z}^N, \text{OPT})\}$; the proposed solution method will later add additional elements to this set. The following optimization problem yields a scenario $\bar{\boldsymbol{z}}$ where $\hat{\boldsymbol{x}}$ is most suboptimal, and provides the ARO Stage-1 decision $\bar{\boldsymbol{x}}$ that allows for the largest improvement:

$$p(\hat{\boldsymbol{x}}, V) = \min_{\substack{\bar{\boldsymbol{z}}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}, \\ \bar{\boldsymbol{y}}^{\boldsymbol{1}}, \ldots, \bar{\boldsymbol{y}}^{|\boldsymbol{V}|}}} \max_{\hat{\boldsymbol{y}}: \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}\hat{\boldsymbol{y}} \leq \boldsymbol{r}(\bar{\boldsymbol{z}})} (\boldsymbol{c}^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}) - (\boldsymbol{c}^\top \hat{\boldsymbol{x}} + \boldsymbol{d}^\top \hat{\boldsymbol{y}}), \tag{17a}$$

$$\text{s.t.} \quad \boldsymbol{A}\bar{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}} \leq \boldsymbol{r}(\bar{\boldsymbol{z}}), \tag{17b}$$

$$\boldsymbol{A}\bar{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}}^{\boldsymbol{i}} \leq \boldsymbol{r}(\boldsymbol{z}^{\boldsymbol{i}}), \quad \forall (\boldsymbol{z}^{\boldsymbol{i}}, v^i) \in V, \tag{17c}$$

$$\boldsymbol{c}^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}^{\boldsymbol{i}} \leq v^i, \quad \forall (\boldsymbol{z}^{\boldsymbol{i}}, v^i) \in V, \tag{17d}$$

$$\bar{\boldsymbol{z}} \in U. \tag{17e}$$

Constraint (17b) ensures that the Stage-1 decision $\bar{\boldsymbol{x}}$ and the decision $\bar{\boldsymbol{y}}$ for scenario $\bar{\boldsymbol{z}}$ are feasible for that scenario. Due to the current choice of set V, constraint (17c) ensures robust feasibility of $\bar{\boldsymbol{x}}$ and constraint (17d) ensures robust optimality of $\bar{\boldsymbol{x}}$. Hence, $\bar{\boldsymbol{x}}$ is ARO and performs strictly better than $\hat{\boldsymbol{x}}$ on the scenario $\bar{\boldsymbol{z}}$, if the optimal objective value of (17) is strictly negative.

One may note that solving (17) is non-trivial. We first describe how problem (17) can be incorporated in an algorithm that guarantees a PARO solution. Subsequently, we describe an approach to approximately solve (17).

Solution $\bar{\boldsymbol{x}}$ need not be PARO; further improvements may be possible. We propose an algorithm that solves problem (17) multiple times. In each iteration the starting Stage-1 solution is the optimal Stage-1 solution of the previous iteration. Additionally, the scenario where the maximum difference is attained is added to the scenario set V, together with the attained objective value in that scenario. Algorithm 1 describes the algorithm and Lemma 9 proves that it yields a PARO Stage-1 solution.

Algorithm 1: Iterative improvement algorithm

```
begin
    Set k = 0, p_0 = −∞ and V_0 = {(z^1, OPT), ..., (z^N, OPT)};
    Compute ARO solution x^0;
    while p_k < 0 do
        Set k = k + 1;
        if k > 1 then
            Set V_k ← V_{k−1} ∪ {(z^{k−1}, c^⊤ x^{k−1} + d^⊤ y^{k−1})};
        end
        Compute p_k := p(x^{k−1}, V_k) and denote the solution by
          (z^k, x^k, y^k, y^{1,k}, ..., y^{N,k});
    end
    Set x̄ = x^k;
end
```

Lemma 9. *A solution $\bar{\boldsymbol{x}}$ obtained from Algorithm 1 is PARO to* (3).

Proof. See Appendix A.14. □

Algorithm 1 requires solving (17) multiple times, but unfortunately it is intractable in general. The reason is that for the original Stage-1 decision $\hat{\boldsymbol{x}}$, the optimal recourse decision $\hat{\boldsymbol{y}}$ for scenario $\bar{\boldsymbol{z}}$ needs to be chosen adversely. However, the set of feasible recourse decisions depends on the scenario $\bar{\boldsymbol{z}}$. Dualizing the inner maximization problem yields

$$\begin{aligned}
\min_{\substack{\bar{\boldsymbol{z}},\bar{\boldsymbol{x}},\bar{\boldsymbol{y}},\boldsymbol{\lambda}\\ \bar{\boldsymbol{y}}^{\boldsymbol{1}},\dots,\bar{\boldsymbol{y}}^{|\boldsymbol{V}|}}} \quad & (\boldsymbol{c}^\top\bar{\boldsymbol{x}} + \boldsymbol{d}^\top\bar{\boldsymbol{y}}) - (\boldsymbol{c}^\top\hat{\boldsymbol{x}} + \boldsymbol{\lambda}^\top(\boldsymbol{r}(\bar{\boldsymbol{z}}) - \boldsymbol{A}\hat{\boldsymbol{x}})),\\
\text{s.t.} \quad & \boldsymbol{A}\bar{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}} \leq \boldsymbol{r}(\bar{\boldsymbol{z}}),\\
& \boldsymbol{A}\bar{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}}^{\boldsymbol{i}} \leq \boldsymbol{r}(\boldsymbol{z}^{\boldsymbol{i}}), \quad \forall(\boldsymbol{z}^{\boldsymbol{i}}, v^i) \in V,\\
& \boldsymbol{c}^\top\bar{\boldsymbol{x}} + \boldsymbol{d}^\top\bar{\boldsymbol{y}}^{\boldsymbol{i}} \leq v^i, \quad \forall(\boldsymbol{z}^{\boldsymbol{i}}, v^i) \in V,\\
& \bar{\boldsymbol{z}} \in U,\\
& \boldsymbol{\lambda}^\top\boldsymbol{B} = \boldsymbol{d}, \quad \boldsymbol{\lambda} \leq \boldsymbol{0}.
\end{aligned}$$

The objective contains bilinear terms. We propose to use a simple alternating direction heuristic, also known as mountain climbing, which guarantees a local optimum (Konno, 1976). For some initial $\bar{\boldsymbol{z}}$ one can determine the optimal $\{\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}, \bar{\boldsymbol{y}}^{\boldsymbol{1}}, \dots, \bar{\boldsymbol{y}}^{|\boldsymbol{V}|}\}$ by solving an LP. Subsequently, we alternate between optimizing for $\boldsymbol{\lambda}$ and $\{\bar{\boldsymbol{z}}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}, \bar{\boldsymbol{y}}^{\boldsymbol{1}}, \dots, \bar{\boldsymbol{y}}^{|\boldsymbol{V}|}\}$ while keeping the other set of variables at their current value. For either set of variables, the problem is an LP. This is continued until two consecutive LP problems yield the same objective value.

The solution quality of Algorithm 1 depends on the starting $\bar{\boldsymbol{z}}$. One option is to simply pick the nominal scenario, if it is defined. An alternative starting solution can be obtained by plugging in an LDR for $\hat{\boldsymbol{y}}$ in (17), i.e., solving

$$\max_{\boldsymbol{w},\boldsymbol{W}} \min_{\substack{\bar{\boldsymbol{z}},\bar{\boldsymbol{x}},\bar{\boldsymbol{y}},\\ \bar{\boldsymbol{y}}^{\boldsymbol{1}},\dots,\bar{\boldsymbol{y}}^{|\boldsymbol{V}|}}} \Big\{(\boldsymbol{c}^\top\bar{\boldsymbol{x}} + \boldsymbol{d}^\top\bar{\boldsymbol{y}}) - (\boldsymbol{c}^\top\hat{\boldsymbol{x}} + \boldsymbol{d}^\top(\boldsymbol{w} + \boldsymbol{W}\bar{\boldsymbol{z}})) \mid \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}(\boldsymbol{w} + \boldsymbol{W}\bar{\boldsymbol{z}}) \leq \boldsymbol{r}(\bar{\boldsymbol{z}})\Big\},$$

additionally subject to (17b)-(17e). This is a static linear robust optimization problem.

The current algorithm may be improved and/or generalized in several ways. One can use different partitions of the set of variables; the currently presented partition works best in our numerical experiments. Also, instead of the alternating direction/mountain climbing heuristic, more advanced methods can be used to solve the resulting bilinear problems (Konno, 1976; Nahapetyan, 2009; Zhen et al., 2018b). Lastly, the presented bilinear approach can also be applied to problems with uncertain $\boldsymbol{A}(\boldsymbol{z})$ and/or $\boldsymbol{c}(\boldsymbol{z})$. This gives rise to bilinear terms in both the objective and constraints; it would be interesting to investigate the numerical performance of the currently presented algorithm for those cases.

6 Numerical Experiments

To demonstrate the value of PARO solutions in practice, we focus on two example problems in which (adaptive) robust optimization has been successfully applied: an inventory management problem and a facility location problem.

For both examples, we will study formulations that have right-hand side uncertainty, and consider instances that are small enough so that the vertices of the uncertainty set can be enumerated. Thus, we can obtain an ARO solution $\boldsymbol{x}_{\text{ARO}}$ by defining a separate recourse variable for each vertex of the uncertainty set. Moreover, Algorithm 1 of Section 5.4 can be used; denote the approximate PARO solution by $\boldsymbol{x}_{\text{PARO}}$.

For comparison purposes, we also compute a PRO solution to (3) using the methodology of Iancu and Trichakis (2014, Theorem 1), which we repeat for convenience. Specifically, we plug in LDR $\boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{w} + \boldsymbol{W}\boldsymbol{z}$, and obtain solution $(\boldsymbol{x}_1, \boldsymbol{w}_1, \boldsymbol{W}_1)$. Subsequently, we optimize for the nominal scenario $\bar{\boldsymbol{z}}$ whilst ensuring that performance in other scenarios does not deteriorate, and feasibility is maintained:

$$\min_{\substack{\boldsymbol{x},\boldsymbol{w},\boldsymbol{W}\\ \boldsymbol{x_2},\boldsymbol{w_2},\boldsymbol{W_2}}} \quad \boldsymbol{c}(\bar{\boldsymbol{z}})^\top \boldsymbol{x_2} + \boldsymbol{d}^\top(\boldsymbol{w_2} + \boldsymbol{W_2}\bar{\boldsymbol{z}}), \tag{18a}$$

$$\text{s.t.} \quad \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x_2} + \boldsymbol{d}^\top(\boldsymbol{w_2} + \boldsymbol{W_2}\boldsymbol{z}) \leq 0, \quad \forall \boldsymbol{z} \in U, \tag{18b}$$

$$\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{w} + \boldsymbol{W}\boldsymbol{z}) \leq \boldsymbol{r}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U, \tag{18c}$$

$$\boldsymbol{x} = \boldsymbol{x_1} + \boldsymbol{x_2}, \boldsymbol{w} = \boldsymbol{w_1} + \boldsymbol{w_2}, \boldsymbol{W} = \boldsymbol{W_1} + \boldsymbol{W_2}. \tag{18d}$$

Constraint (18d) states that the new solution equals the original solution (variables with subscript 1) plus an adaptation (variables with subscript 2). Constraint (18b) ensures that the adaptation does not deteriorate performance in any scenario, and the objective is to optimize performance in scenario $\bar{\boldsymbol{z}}$. According to Iancu and Trichakis (2014, Theorem 1), the optimal solution for $(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{W})$ is PRO to (3). Let $\boldsymbol{x}_{\text{PRO}}$ denote the optimal solution for $\boldsymbol{x}$.

We compare the performance of the three Stage-1 (here-and-now) solutions $\boldsymbol{x}_{\text{ARO}}$, $\boldsymbol{x}_{\text{PARO}}$ and $\boldsymbol{x}_{\text{PRO}}$. For solutions $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PARO}}$ we use the optimal recourse decision, for $\boldsymbol{x}_{\text{PRO}}$ recourse decisions are computed using the LDR. Let $v(\boldsymbol{x}, \boldsymbol{z})$ denote the objective value in scenario $\boldsymbol{z}$ resulting from solution $\boldsymbol{x}$ and its corresponding decision rule. We consider the relative improvement in objective value of $\boldsymbol{x}_{\text{PARO}}$ w.r.t. $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PRO}}$. In case of maximization we compute

$$\frac{v(\boldsymbol{x}_{\text{PARO}}, \boldsymbol{z}) - v(\boldsymbol{x}_{\text{ARO}}, \boldsymbol{z})}{v(\boldsymbol{x}_{\text{ARO}}, \boldsymbol{z})} \cdot 100\% \text{ and } \frac{v(\boldsymbol{x}_{\text{PARO}}, \boldsymbol{z}) - v(\boldsymbol{x}_{\text{PRO}}, \boldsymbol{z})}{v(\boldsymbol{x}_{\text{PRO}}, \boldsymbol{z})} \cdot 100\%, \tag{19}$$

and in case of minimization we multiply by -1. We report the following three measures:

Nominal: Relative improvement in nominal scenario $\bar{\boldsymbol{z}}$.

Average: Average relative improvement over 10 uniform randomly sampled scenarios.

Maximum: Relative improvement in the scenario with the maximum performance difference between $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PARO}}$. This scenario, which we denote $\boldsymbol{z}^*$, is found by solving (17) with fixed $\hat{\boldsymbol{x}} = \boldsymbol{x}_{\text{ARO}}$ and $\bar{\boldsymbol{x}} = \boldsymbol{x}_{\text{PARO}}$.

All optimization problems are solved using Gurobi 9.0 (Gurobi Optimization LLC, 2020) with the dual simplex algorithm selected. We note that the influence of different solvers may also be investigated, but this is beyond the scope of this paper.

During our numerical studies we found examples where Algorithm 1 was not able to improve upon the initial Stage-1 solution $\boldsymbol{x}_{\text{ARO}}$. This could occur if the initial $\boldsymbol{x}_{\text{ARO}}$ happens to be PARO. Or, it could occur if there is a unique ARO solution - after all, not every ARO instance has multiple worst-case optimal Stage-1 solutions. The latter has been reported before in literature. De Ruiter et al. (2016) show that the multi-stage production-inventory model of Ben-Tal et al. (2004) has unique here-and-now decisions in almost all time periods, if LDRs are used. In that example, the reported multiplicity of solutions is mainly due to non-PRO decision rule coefficients. We find that multiplicity of Stage-1 solutions appears in particular when problem data is integer.

6.1 Inventory management

Problem description

Iancu and Trichakis (2014, Section 5.2) considers a two-stage inventory management example with a single warehouse and n retailers. Demand d_i at each retail location is unknown at the moment that stocking decisions need to be made. The sales for a location i depend on the stock allocated to the location and the demand d_i at the location, $i = 1, \dots, n$.[2] The optimal sales at location i are equal to the minimum of the available stock and the demand at that location. Thus, once the stocking decision has been made and the demand has realized, the optimal sales are readily calculated. But, the Stage-2 decision variable is solely an auxiliary variable and does not represent an actual managerial decision.

We extend the example to a genuine two-stage problem, where the Stage-2 decision variables represent actual decisions. The demand in retail location i is given by

$$d_i = d_i^0 + \boldsymbol{q}_i^\top \boldsymbol{z},$$

where d_i^0 is the nominal demand, $\boldsymbol{z} \in \mathbb{R}^{n_f}$ are market factors that influence the demand, and $\boldsymbol{q_i}$ indicates how demand in store i is influenced by the market factors. Market factors $\boldsymbol{z}$ are the primitive uncertainties and reside in uncertainty set $U := \{\boldsymbol{z} : -b{\cdot}\mathbf{1} \leq \boldsymbol{z} \leq b{\cdot}\mathbf{1},\ -B \leq \mathbf{1}^\top \boldsymbol{z} \leq B\}$.

Let $\boldsymbol{x} \in \mathbb{R}^n$ denote the stocking decisions. The warehouse has C units available and each store i has capacity c_i. Store i has holding costs h_i and transport from the warehouse to this store costs $t_{0,i}$ per unit. After market factors $\boldsymbol{z}$ are revealed, $\alpha \cdot 100\%$ of initial stock can be quickly re-allocated before the demand actually occurs. Adaptive variable $u_{i,j}(\boldsymbol{z})$ represents stock re-allocated from store i to store j; the associated cost is t_{ij}. The sales in store i are represented by an adaptive variable $y_i(\boldsymbol{z})$ with revenue r_i per sold unit.

Stocking and re-allocation decisions are to be made to maximize worst-case profit by solving:

$$\max_{\boldsymbol{x},\boldsymbol{y}(\cdot),\boldsymbol{u}(\cdot)} \quad \min_{\boldsymbol{z}\in U} \sum_{i=1}^{n} r_i y_i(\boldsymbol{z}) - \sum_{i=1}^{n} (t_{0i} + h_i)^\top x_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} t_{ij} |u_{ij}(\boldsymbol{z})|, \tag{20a}$$

$$\text{s.t.} \quad y_i(\boldsymbol{z}) \leq x_i - \sum_{j>i} u_{ij}(\boldsymbol{z}) + \sum_{j<i} u_{ji}(\boldsymbol{z}), \quad \forall i\ \forall \boldsymbol{z} \in U, \tag{20b}$$

$$y_i(\boldsymbol{z}) \leq d_i^0 + \boldsymbol{q}_i^\top \boldsymbol{z}, \quad \forall i\ \forall \boldsymbol{z} \in U \tag{20c}$$

$$y_i(\boldsymbol{z}) \geq 0, \quad \forall i\ \forall \boldsymbol{z} \in U, \tag{20d}$$

$$\sum_{i=1}^{n} x_i = C, \tag{20e}$$

$$0 \leq x_i \leq c_i, \quad \forall i, \tag{20f}$$

$$\sum_{i=1}^{n} \sum_{j>i} |u_{ij}(\boldsymbol{z})| \leq \alpha C, \quad \forall \boldsymbol{z} \in U, \tag{20g}$$

$$u_{ij} = -u_{ji}, \ \forall i, j, \ \forall \boldsymbol{z} \in U. \tag{20h}$$

The last term in the objective subtracts transportation cost t_{ij} per item moved from i to j. The associated absolute value terms can be linearized using additional variables. Constraint (20b) ensures that the sales at point i do not exceed the inventory at point i after the re-allocation of stock. Constraint (20g) ensures that at most $\alpha \cdot 100\%$ of stock is re-allocated. Constraint (20h) ensures that sales at each location are nonnegative.

Data

We consider 1000 randomly generated instances. Where applicable, parameters are sampled and/or chosen equal to Iancu and Trichakis (2014); choices are repeated for convenience. We set instance size $n = 10$ and available inventory $C = 2000$. We use $\alpha = 0.05$, i.e., at most 5% of inventory may be re-allocated. All other parameters are independently drawn from a discrete uniform distribution. Individual capacities $\boldsymbol{c} \in \mathbb{R}^n$ are drawn between 300 and 500, operating

[2] In essence, this is a single-stage problem. However, sales at each location are usually modeled as adaptive variables, yielding a two-stage ARO problem.

cost $\boldsymbol{h} \in \mathbb{R}^n$ are drawn between 1 and 3, revenue rates $\boldsymbol{r} \in \mathbb{R}^n$ are drawn between 20 and 40 and nominal demand $\boldsymbol{d}_0$ is drawn between 100 and 200. Retail locations are uniformly distributed on a 7×7 grid with the warehouse at the center. Transportation cost t_{ij} between two retail locations i and j is the Euclidean distance rounded to the nearest integer.

The number of market factors n_f is drawn from values 2, 3 and 4, and exposure parameters $\boldsymbol{q_i} \in \mathbb{R}^{n_f}$ are drawn from a continuous uniform distribution between -2 and 2 for each i. The uncertainty set for the factor values is described by $b = 5$ and $B = 25$. The nominal scenario is the scenario without shocks, i.e., $\bar{z}_j = 0$ for all j. Note that $\bar{\boldsymbol{z}} \in \text{ri}(U)$.

Results

The worst-case objective value of $\boldsymbol{x}_{\text{PRO}}$ is within 0.24% of the worst-case objective value of $\boldsymbol{x}_{\text{ARO}}$ for all instances. In 80% of the instances the solutions $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PARO}}$ differ. Table 1 reports the median and maximum difference in ℓ_1-norm for these instances, multiplied by $\frac{1}{2}$. This can be interpreted as the total difference in initial stock allocation. Total allocated stock equals $C = 2000$, so the different solutions allocate up to a quarter of stock differently.

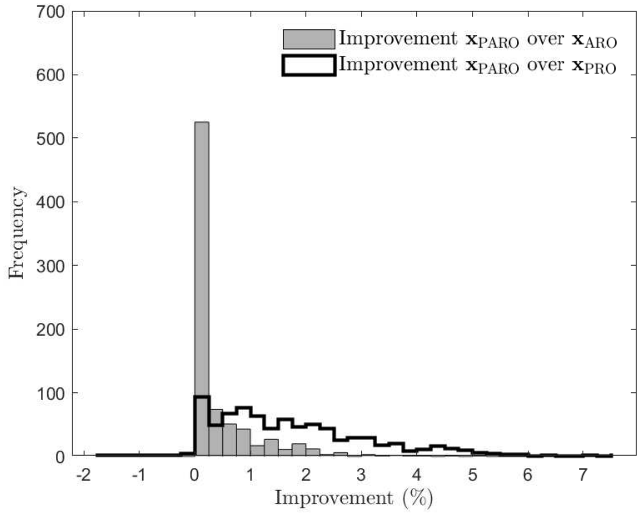
	$\frac{1}{2}\|\boldsymbol{x}_{\text{PARO}} - \boldsymbol{x}_{\text{ARO}}\|_1$	$\frac{1}{2}\|\boldsymbol{x}_{\text{PARO}} - \boldsymbol{x}_{\text{PRO}}\|_1$	$\frac{1}{2}\|\boldsymbol{x}_{\text{ARO}} - \boldsymbol{x}_{\text{PRO}}\|_1$
median	83	64	89
max	453	494	517

Table 1: Total differences in Stage-1 stock allocations.

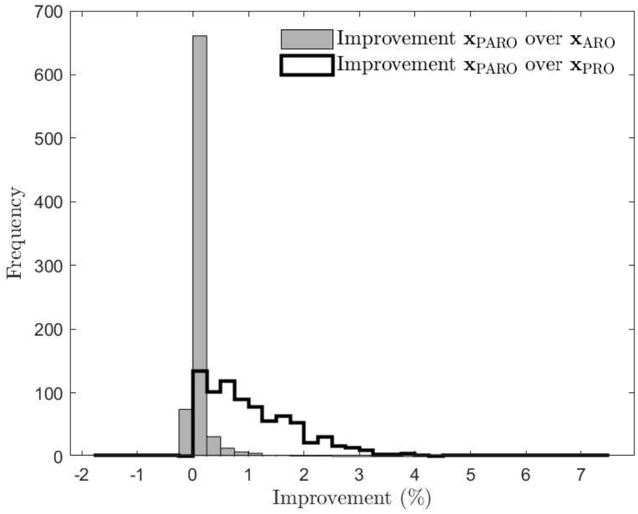
Figure 2 shows histograms of the relative objective value improvement of $\boldsymbol{x}_{\text{PARO}}$ over $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PRO}}$ (according to (19)) for the instances with different $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PARO}}$. Figure 2a shows the improvement for maximum difference scenario $\boldsymbol{z}^*$. The median improvements w.r.t. $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PRO}}$ are 0.06% and 1.5%. The maximum improvements are 6.9% and 7.3%, respectively. Figure 2b shows the improvement for nominal scenario $\boldsymbol{z}$. The median improvements w.r.t. $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PRO}}$ are $< 0.01\%$ and 0.89%, respectively. The maximum improvements are 2.5% and 4.3%. This indicates that, even though $\boldsymbol{x}_{\text{PRO}}$ optimizes for the nominal scenario, further improvements can be obtained by using optimal decision rules. Average improvements for 10 random scenarios in the uncertainty set are shown in Figure 2c. The median improvements w.r.t. $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PRO}}$ are $< 0.01\%$ and 0.72%, respectively. The maximum improvements are 2.2% and 4.6%.

The median improvement of $\boldsymbol{x}_{\text{PARO}}$ over $\boldsymbol{x}_{\text{ARO}}$ is small in magnitude, for all three measures. However, the maximum improvement is several percentage points, which is noteworthy. If the Stage-1 solution represents a decision that is to be implemented in practice, the possibility to get an improvement of up to several percentage points warrants the extra effort to obtain an (approximate) PARO solution. Moreover, we compare $\boldsymbol{x}_{\text{PARO}}$ to the *first* ARO solution $\boldsymbol{x}_{\text{ARO}}$ that is found using Gurobi. It is possible that there exists yet another ARO solution, for which the improvement percentages are larger than those reported in Figure 2.

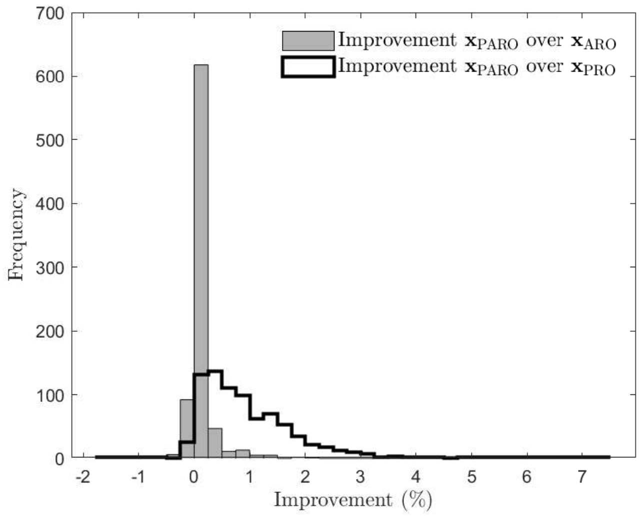
The improvement of $\boldsymbol{x}_{\text{PARO}}$ over $\boldsymbol{x}_{\text{PRO}}$ is larger than that of $\boldsymbol{x}_{\text{PARO}}$ over $\boldsymbol{x}_{\text{ARO}}$ in all three figures. This indicates that $\boldsymbol{x}_{\text{ARO}}$ performs better than $\boldsymbol{x}_{\text{PRO}}$ on all three measures. This is particularly noteworthy for the performance in nominal scenario $\bar{\boldsymbol{z}}$. Solution $\boldsymbol{x}_{\text{PRO}}$ need not be ARO in theory, but the results show that it is within 0.24% of worst-case optimality. Moreover, in its second step the PRO methodology optimizes for the nominal scenario $\bar{\boldsymbol{z}}$. Yet, Figure 2b shows that $\boldsymbol{x}_{\text{ARO}}$ performs better on the nominal scenario than $\boldsymbol{x}_{\text{PRO}}$. Thus, whereas the restriction to LDRs does not lead to bad performance in worst-case, it does impact performance in non-worst-case scenarios.



(a) Scenario $\boldsymbol{z}^*$



(b) Scenario $\bar{\boldsymbol{z}}$



(c) 10 Random scenarios

Figure 2: Results for the inventory management example.

6.2 Facility location

Problem description

Consider a strategic decision-making problem where a number of facilities are to be opened, in order to satisfy the demand of a number of customers. The goal is to choose the locations for opening a facility such that the cost opening the facilities plus the transportation cost for satisfying demand is minimized. We consider this problem in a two-stage setting with uncertain demand. Thus, facility opening decisions need to be made in Stage 1, before Stage-2 demand is known.

Suppose there are n locations where a facility can be opened, and m demand locations. Let $\boldsymbol{x} \in \{0,1\}^n$ be a binary Stage-1 decision variable denoting the facility opening decisions. Opening facility costs f_i and yields a capacity s_i, $i = 1, \dots, n$. Let $\boldsymbol{y} \in \mathcal{R}^{m,m\times n}$ be the Stage-2 decision variable denoting transport from facility i to demand location j; let c_{ij} denote the associated costs, $i = 1, \dots, n$, $j = 1, \dots, m$. Let z_j denote the uncertain demand in location j.

The two-stage facility location model reads

$$\min_{\boldsymbol{x},\boldsymbol{y}(\cdot)} \quad \max_{\boldsymbol{z}\in U} \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} y_{ij}(\boldsymbol{z}) + \sum_{i=1}^{n} f_i x_i, \tag{21a}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} y_{ij}(\boldsymbol{z}) \geq z_j, \quad \forall \boldsymbol{z} \in U,\ \forall j = 1,\dots,m, \tag{21b}$$

$$\sum_{j=1}^{m} y_{ij}(\boldsymbol{z}) \leq s_i x_i, \quad \forall \boldsymbol{z} \in U,\ \forall i = 1,\dots,m, \tag{21c}$$

$$y_{ij}(\boldsymbol{z}) \geq 0,\ \forall \boldsymbol{z} \in U, \quad \forall i = 1,\dots,n,\ j = 1,\dots,m, \tag{21d}$$

$$\boldsymbol{x} \in \{0,1\}^n, \tag{21e}$$

with uncertainty set

$$U = \{\boldsymbol{z} : \sum_{j=1}^{m} z_j \leq \Gamma,\ l_j \leq z_j \leq u_j,\ \forall j = 1,\dots,m\}.$$

Data

We consider 1000 instances with $m = 8$ demand locations and $n = 20$ possible facility locations. Facility capacity s_i is set at 15 for each i. Other parameters are independently drawn from a discrete uniform distribution. Construction costs $\boldsymbol{f} \in \mathbb{R}^n$ are drawn between 4 and 22. Entries of transportation cost matrix $\boldsymbol{C} \in \mathbb{R}^{n\times m}$ are drawn between 2 and 12.

We set lower and upper bound $l_j = 8$ and $u_j = 12$ for each demand location $j = 1,\dots,m$, and set maximum total demand $\Gamma = 90$. The nominal demand scenario is $\bar{z}_j = 10$ for all j. Note that $\bar{\boldsymbol{z}} \in \text{ri}(U)$.

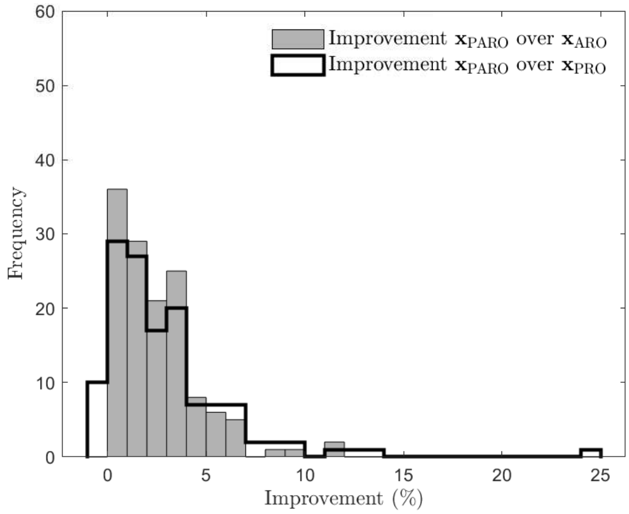
Results

The worst-case objective value of $\boldsymbol{x}_{\text{PRO}}$ is within 0.72% of the worst-case objective value of $\boldsymbol{x}_{\text{ARO}}$ for all instances. In 13% of the instances the solutions $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PARO}}$ differ. Table 2 reports the median and maximum difference in ℓ_1-norm for these instances. This represents the number of different facilities that are opened. For example, an ℓ_1-norm of 2 indicates that one solution opened facility i and another solution opened facility j, or one solution opened both facilities i and j and the other solution opened neither. The total number of considered facility locations is $n = 20$, so the differences reported in Table 2 are substantial.

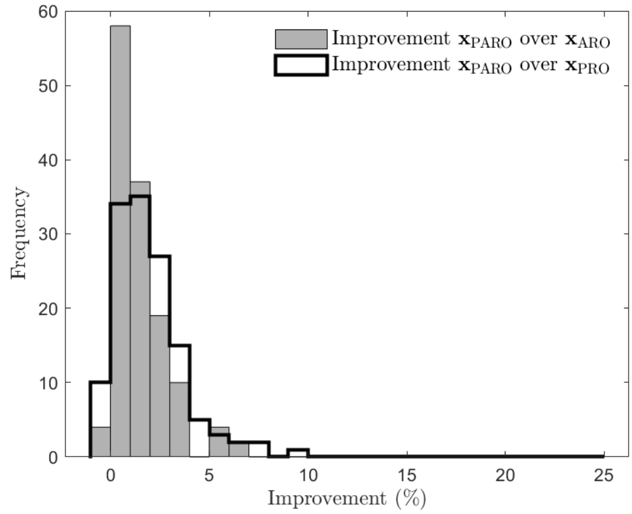
	$\lVert\boldsymbol{x}_{\text{PARO}} - \boldsymbol{x}_{\text{ARO}}\rVert_1$	$\lVert\boldsymbol{x}_{\text{PARO}} - \boldsymbol{x}_{\text{PRO}}\rVert_1$	$\lVert\boldsymbol{x}_{\text{ARO}} - \boldsymbol{x}_{\text{PRO}}\rVert_1$
median	2	2	0
max	7	8	9

Table 2: Total differences in Stage-1 facility openings.

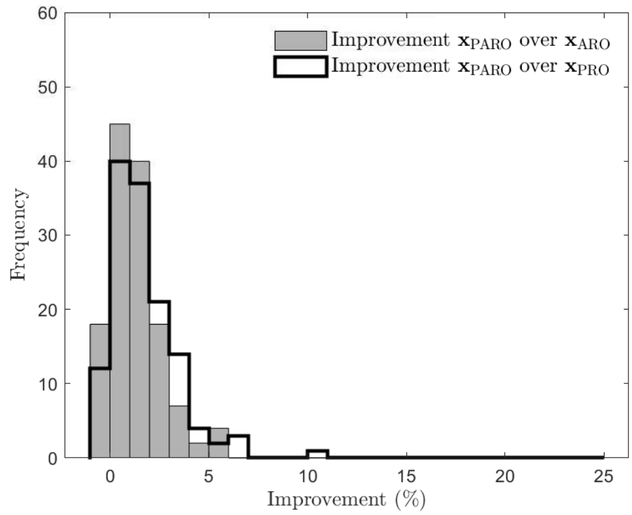
Figure 3 shows histograms of the relative objective value improvement of $\boldsymbol{x}_{\text{PARO}}$ over $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PRO}}$ (according to (19)) for the 14% of instances where $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PARO}}$ differ. Figure 3a show the improvement for maximum difference scenario $\boldsymbol{z}^*$. The median improvement w.r.t. both $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PRO}}$ is 2.1%, with the maximum improvements 11% and 24%, respectively. Figure 3b show the improvement for nominal scenario $\boldsymbol{z}$. The median improvements w.r.t. $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PRO}}$ are 1.3% and 1.5%, respectively. The maximum improvements are 6.4% and 10%. Figure 3c show the improvement for 10 random scenarios in the uncertainty set. The median improvements w.r.t. $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PRO}}$ are again 1.2% and 1.5%, respectively. The maximum improvements are 5.5% and 11%.



(a) Scenario $\boldsymbol{z}^*$



(b) Scenario $\bar{\boldsymbol{z}}$



(c) 10 Random scenarios

Figure 3: Results for the facility location example.

Compared to the inventory management example of Section 6.1, the percentage of instances with different $\boldsymbol{x}_{\text{ARO}}$ and $\boldsymbol{x}_{\text{PARO}}$ is lower, but the magnitude of improvements is larger. Similar to Section 6.1, we note that it is possible that there exists yet another ARO solution, for which the improvement percentages are larger than those reported in Figure 3.

The improvement of $\boldsymbol{x}_{\text{PARO}}$ over $\boldsymbol{x}_{\text{PRO}}$ is similar to that of $\boldsymbol{x}_{\text{PARO}}$ over $\boldsymbol{x}_{\text{ARO}}$ in all three figures. The advantage of $\boldsymbol{x}_{\text{ARO}}$ is that it is worst-case optimal, and uses the optimal recourse decision. However, when planning its Stage-1 decision (the facility locations) it did not account for performance in non-worst-case scenarios. Solution $\boldsymbol{x}_{\text{PRO}}$ is close to worst-case optimal (within 0.72%) and in its second step the PRO methodology optimizes for the nominal scenario $\bar{\boldsymbol{z}}$. However, in Stage 2 it is limited to LDRs. The results in Figure 3 indicate that for all three measures these effects balance each other out.

7 Conclusion

In this paper, we dealt with Pareto efficiency in two-stage adaptive robust optimization problems. Similar to static robust optimization, the large majority of solution techniques focus only on worst-case optimality, and may yield solutions that are not Pareto efficient. To alleviate this, we introduced the concept of Pareto Adaptive Robustly Optimal (PARO) solutions.[3]

Using FME as the predominant technique, we have analyzed the relation between PRO and PARO and investigated optimality of various decision rule structures in both worst-case and non-worst-case scenarios. We have shown the existence of PARO here-and-now decisions and shown that there exists a PWL decision rule that is PARO.

Moreover, we have provided several practical approaches to generate or approximate PARO solutions. Numerical experiments on an inventory management example and a facility location example demonstrate that PARO solutions can significantly improve performance in non-worst-case scenarios over ARO and PRO solutions.

A potential direction for future research would be to further investigate constructive approaches to find or approximate PARO solutions. In particular, it would be valuable to have tractable algorithms for larger instances and/or more general classes of problems than the ones that can be tackled using the currently presented approaches.

References

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

Ben-Tal, A., El Housni, O., and Goyal, V. (2020). A tractable approach for designing piecewise affine policies in two-stage adjustable robust optimization. *Math. Program.*, 182:57–102.

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

Bertsimas, D., Bidkhori, H., and Dunning, I. (2019). The price of flexibility. Unpublished manuscript.

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

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

Bertsimas, D. and Tsitsiklis, J. N. (1997). *Introduction to linear optimization*. Athena Scientific.

De Ruiter, F. J. C. T., Brekelmans, R. C. M., and den Hertog, D. (2016). The impact of the existence of multiple adjustable robust solutions. *Math. Program.*, 160:531–545.

Fourier, J. (1827). Reported in: Analyse des travaux de l'Académie, Royale des Sciences, pendant l'année 1824, Partie mathématique. In *Mémoires de l'Académie des sciences de l'Institut de France*, volume 7. Académie des sciences.

Gabrel, V., Murat, C., and Thiele, A. (2014). Recent advances in robust optimization: an overview. *Eur. J. Oper. Res.*, 235:471–483.

Gorissen, B. L., Yanıkoğlu, İ., and den Hertog, D. (2015). A practical guide to robust optimization. *Omega*, 53:124–137.

Gurobi Optimization LLC (2020). Gurobi optimizer reference manual. `http://www.gurobi.com`.

[3]For ARO problems, every non-PARO solution is dominated by a PARO solution, even if the former is Pareto Robustly Optimal (PRO), as defined by Iancu and Trichakis (2014).

Guslitser, E. (2002). Uncertainty-immunized solutions in linear programming. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.623.2382&rep=rep1&type=pdf. Master's thesis, Technion - Israel Institute of Technology.

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

Iancu, D. A. and Trichakis, N. (2014). Pareto efficiency in robust optimization. *Manag. Sci.*, 60(1):130–147.

Jabr, R. A. (2013). Adjustable robust OPF with renewable energy sources. *IEEE T. Power Syst.*, 28(4):4742–4751.

Konno, H. (1976). A cutting plane algorithm for solving bilinear programs. *Math. Program.*, 11:14–27.

Kuhn, D., Wiesemann, W., and Georghiou, A. (2009). Primal and dual linear decision rules in stochastic and robust optimization. *Math. Program.*, 130:177–209.

Marandi, A. and den Hertog, D. (2018). When are static and adjustable robust optimization problems with constraint-wise uncertainty equivalent? *Math. Program.*, 170(2):555–568.

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

Nahapetyan, A. H. (2009). Bilinear programming. In Floudas, C. A. and Pardalos, P. M., editors, *Encyclopedia of optimization*, pages 279–288. Springer.

Polinder, G. J., Breugem, T., Dollevoet, T., and Maróti, G. (2019). An adjustable robust optimization approach for periodic timetabling. *Transport. Res. B-Meth.*, 128:50–68.

Rockafellar, T. R. (1970). *Convex analysis.* Princeton University Press, Princeton, NJ.

Ten Eikelder, S. C. M., Ajdari, A., Bortfeld, T., and den Hertog, D. (2019). Adjustable robust treatment-length optimization in radiation therapy. arXiv preprint arXiv:1906.12116.

Yanıkoğlu, İ., Gorissen, B. L., and den Hertog, D. (2019). A survey of adjustable robust optimization. *Eur. J. Oper. Res.*, 277:799–813.

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

Zhen, J., den Hertog, D., and Sim, M. (2018a). Adjustable robust optimization via Fourier-Motzkin elimination. *Oper. Res.*, 66(4):1086–1100.

Zhen, J., Marandi, A., den Hertog, D., and Vandenberghe, L. (2018b). Disjoint bilinear programming: A two-stage robust optimization perspective. Available at Optimization Online.

A Technical Lemmas and Proofs

A.1 Bounds on eliminated adaptive variables

Lemma 10. *Let $\boldsymbol{x}$ be ARF to (3). Let $\varphi_i(\boldsymbol{x},\boldsymbol{z}) = \boldsymbol{r}_i(\boldsymbol{z}) - \boldsymbol{a}_i(\boldsymbol{z})^\top \boldsymbol{x}$ for each constraint $i = 1,\dots,m$ of (3b). Consider the system of inequalities $\boldsymbol{b}_i^\top \boldsymbol{y}(\boldsymbol{z}) \leq \varphi_i(\boldsymbol{x},\boldsymbol{z})$, $i = 1,\dots,m$ and use FME to eliminate all variables. For all $k = 1,\dots,n_y$ we can write the bounds after elimination of variable $y_k(\boldsymbol{z})$ as*

$$\begin{aligned}\max_{S_k \in C_k^-} \Big\{ \sum_{p \in S_k} \alpha(S_k,p)\varphi_p(\boldsymbol{x},\boldsymbol{z}) - \sum_{l=k+1}^{n_y} \beta(S_k,l) y_l(\boldsymbol{z}) \Big\} &\leq y_k(\boldsymbol{z}) \\ &\leq \min_{T_k \in C_k^+} \Big\{ \sum_{q \in T} \alpha(T_k,q)\varphi_q(\boldsymbol{x},\boldsymbol{z}) - \sum_{l=k+1}^{n_y} \beta(T_k,l) y_l(\boldsymbol{z}) \Big\}, \quad \forall \boldsymbol{z} \in U,\end{aligned} \tag{A.1}$$

for some coefficients α and β independent of $\boldsymbol{z}$, and $C_k^-, C_k^+ \subseteq P(\{1,\dots,m\})$, with $P(\{1,\dots,m\})$ the power set of $\{1,\dots,m\}$. Additionally, if $S_k \in C_k^-$ for some k, then $\alpha(S_k,p) < 0$ for all $p \in S_k$. If $T_k \in C_k^+$ for some k, then $\alpha(T_k,q) > 0$ for all $q \in T_k$.

Proof. Proof by induction.

Base case:
Elimination of variable $y_1(\boldsymbol{z})$ yields

$$\max_{\{p:b_{p,1}<0\}} \left\{ \frac{\varphi_p(\boldsymbol{x},\boldsymbol{z})}{b_{p,1}} - \frac{\sum_{l=2}^{n_y} b_{p,l} y_l(\boldsymbol{z})}{b_{p,1}} \right\} \leq y_1(\boldsymbol{z}) \leq \min_{\{q:b_{q,1}>0\}} \left\{ \frac{\varphi_q(\boldsymbol{x},\boldsymbol{z})}{b_{q,1}} - \frac{\sum_{l=2}^{n_y} b_{q,l} y_l(\boldsymbol{z})}{b_{q,1}} \right\}. \tag{A.2}$$

Define

$$C_1^- = \{p \mid b_{p,1} < 0\}, \quad C_1^+ = \{q \mid b_{q,1} > 0\},$$

then each constraint in C_1^- defines a lower bound on $y_1(\boldsymbol{z})$ and each constraint in C_1^+ defines an upper bound on $y_1(\boldsymbol{z})$. Each element of C_1^- and C_1^+ is an individual 'original' constraint index and not a set of constraints indices. For all $S_1 = \{p\} \in C_1^-$ set $\alpha(S,p) = b_{p,1}^{-1}$, and for all $T_1 = \{q\} \in C_1^+$ set $\alpha(T,q) = b_{q,1}^{-1}$. Furthermore, set $\beta(S_1,l) = b_{p,l} B_{p,1}^{-1}$ for all $S_1 = \{p\} \in C_1^- \cup C_1^+$ and all $l = 2,\ldots,n_y$. With these definitions, (A.2) is reformulated in form (A.1). Additionally, by construction, $\alpha(S_1,p) < 0$ if $p \in S_1$, $S_1 \in C_1^-$ and $\alpha(T_1,q) > 0$ if $q \in T_1$, $T_1 \in C_1^+$.

Induction step:
Suppose the result holds for some $k-1$ (i.e., after elimination of variable $y_{k-1}(\boldsymbol{z})$). Variable $y_k(\boldsymbol{z})$ can occur in two types of constraints: (i) original constraints $i = 1,\ldots,m$ that do not depend on $y_1(\boldsymbol{z}),\ldots,y_{k-1}(\boldsymbol{z})$) and (ii) the new constraints acquired after elimination of $y_1(\boldsymbol{z}),\ldots,y_{k-1}(\boldsymbol{z})$). For case (i), define

$$\begin{aligned} I_k^- &= \{p \mid b_{p,k} < 0,\ b_{p,l} = 0,\ \forall l = 1,\ldots,k-1\}, \\ I_k^+ &= \{p \mid b_{p,k} > 0,\ b_{p,l} = 0,\ \forall l = 1,\ldots,k-1\}, \end{aligned}$$

then each constraint in I_k^- defines a lower bound on $y_k(\boldsymbol{z})$ and each constraint in I_k^+ provides an upper bound on $y_k(\boldsymbol{z})$. Reformulation to form (A.1) is similar to the case $k = 1$. Thus, $\alpha(S_k,p) < 0$ if $p \in S_k$, $S_k \in I_k^-$ and $\alpha(T_k,p) > 0$ if $p \in T_k$, $T_k \in I_k^+$.

For case (ii), $y_k(\boldsymbol{z})$ can occur in constraints resulting from picking linear lower and upper bounds on $y_l(\boldsymbol{z})$ from (A.1). If these bounds are independent of $y_{l+1}(\boldsymbol{z}),\ldots,y_{k-1}(\boldsymbol{z})$, for $l = 1,\ldots,k-1$, they are used directly to eliminate $y_k(\boldsymbol{z})$. For any such pair of constraints $S_l \in C_l^-$ and $T_l \in C_l^+$, FME yields the following bound on $y_k(\boldsymbol{z})$ (due to the induction assumption):

$$\begin{aligned} \sum_{p\in S_l} \alpha(S_l,p)\varphi_p(\boldsymbol{x},\boldsymbol{z}) - \sum_{q\in T_l} \alpha(T_l,q)\varphi_q(\boldsymbol{x},\boldsymbol{z}) - \sum_{l=k+1}^{n_y} y_l(\boldsymbol{z})\big(\beta(S_l,l) - \beta(T_l,l)\big) \\ \leq y_k(\boldsymbol{z})\big(\beta(S_l,k) - \beta(T_l,k)\big). \end{aligned} \tag{A.3}$$

We proceed by dividing by the coefficient of $y_k(\boldsymbol{z})$. If $\beta(S_l,k) > \beta(T_l,k)$, inequality (A.3) defines a lower bound for $y_k(\boldsymbol{z})$; if $\beta(S_l,k) < \beta(T_l,k)$, inequality (A.3) defines an upper bound for $y_k(\boldsymbol{z})$. Define

$$\begin{aligned} J_k^- = \{S_k \mid \exists l = 1,\ldots,k-1 \text{ s.t. } & S_k = S_l \cup T_l,\ S_l \in C_l^-,\ T_l \in C_l^+, \\ & \beta(S_l,j) = \beta(T_l,j),\ \forall j < l,\ \beta(S_l,k) > \beta(T_l,k)\}, \\ J_k^+ = \{T_k \mid \exists l = 1,\ldots,k-1 \text{ s.t. } & T_k = S_l \cup T_l,\ S_l \in C_l^-,\ T_l \in C_l^+, \\ & \beta(S_l,j) = \beta(T_l,j),\ \forall j < l,\ \beta(S_l,k) < \beta(T_l,k)\} \end{aligned}$$

so each element S_k in J_k^- (or T_k in J_k^+) is a union of the indices of a lower bound constraint (set S_l) and an upper bound constraint (set T_l) on $y_l(\boldsymbol{z})$. The condition $\beta(S_l,j) = \beta(T_l,j)$, $\forall j < l$ on the second line ensures that these lower and upper bound constraints on $y_l(\boldsymbol{z})$ do not specify a constraint on $y_{l+1}(\boldsymbol{z}),\ldots,y_{k-1}(\boldsymbol{z})$.

Set the coefficients for the not yet eliminated variables $y_{k+1}(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z})$ for form (A.1) as

$$\beta(S_k,j) = \frac{\beta(S_l,j) - \beta(T_l,j)}{\beta(S_l,k) - \beta(T_l,k)},\ \forall j = k+1,\ldots,n_y.$$

If $S_k \in J_k^-$, with $S_k = S_l \cup T_l$ for some $S_l \in C_l^-$ and $T_l \in C_l^+$, $l = 1, \dots, k-1$, then set

$$\alpha(S_k, p) = \begin{cases} \dfrac{\alpha(S_l, p)}{\beta(S_l, k) - \beta(T_l, k)} & \text{if } p \in S_l,\ p \notin T_l, \\ \dfrac{\alpha(S_l, p) - \alpha(T_l, p)}{\beta(S_l, k) - \beta(T_l, k)} & \text{if } p \in S_l \cap T_l, \\ \dfrac{-\alpha(T_l, p)}{\beta(S_l, k) - \beta(T_l, k)} & \text{if } p \notin S_l,\ p \in T_l. \end{cases} \tag{A.4}$$

Similarly, if $T_k \in J_k^+$, with $T_k = S_l \cup T_l$ for some $S_l \in C_l^-$ and $T_l \in C_l^+$ for some $l = 1, \dots, k-1$, then set

$$\alpha(T_k, p) = \begin{cases} \dfrac{\alpha(S_l, p)}{\beta(S_l, k) - \beta(T_l, k)} & \text{if } p \in S_l,\ p \notin T_l, \\ \dfrac{\alpha(S_l, p) - \alpha(T_l, p)}{\beta(S_l, k) - \beta(T_l, k)} & \text{if } p \in S_l, \cap T_l \\ \dfrac{-\alpha(T_l, p)}{\beta(S_l, k) - \beta(T_l, k)} & \text{if } p \notin S_l,\ p \in T_l. \end{cases} \tag{A.5}$$

Due to the induction hypothesis, $\alpha(S_l, p) < 0$ if $S_l \in C_l^-$ and $\alpha(T_l, p) > 0$ if $T_l \in C_l^+$ for $l < k$. The denominator in both lines of (A.4) is positive, so in that case $\alpha(S_k, p) < 0$. The denominator in both lines of (A.5) is negative, so in that case $\alpha(T_k, p) > 0$. With the new coefficients chosen as above, (A.3) provides a lower or upper bound on $y_{\hat{k}}(\boldsymbol{z})$ of the form inside the maximum or minimum operator in (A.1), respectively.

Finally, define $C_k^- = I_k^- \cup J_k^-$ and $C_k^+ = I_k^+ \cup J_k^+$. Each constraint in C_k^- defines a lower bound on $y_k(\boldsymbol{z})$ and each constraint in C_k^+ defines an upper bound on $y_k(\boldsymbol{z})$. Moreover, set $C_k = C_k^- \cup C_k^+$ contains all constraints after elimination of $y_1(\boldsymbol{z}), \dots, y_{k-1}(\boldsymbol{z})$ that have $y_k(\boldsymbol{z})$ as lowest indexed adaptive variable. This completes the induction step. □

A.2 Proof Lemma 1

We consider only adaptive robust feasibility and not optimality, so the objective of P_{hybrid} can be ignored. According to Lemma 10, each adaptive variable $y_k(\boldsymbol{z})$, $k = 1, \dots, n_y$ must satisfy bounds (A.1). For P_{hybrid} term $\varphi_i(\hat{\boldsymbol{z}}, \boldsymbol{z}_{(i)}) = r_i(\hat{\boldsymbol{z}}, \boldsymbol{z}_{(i)}) - \boldsymbol{a}_i(\hat{\boldsymbol{z}}, \boldsymbol{z}_{(i)})^\top \boldsymbol{x}$ depends only on $\hat{\boldsymbol{z}}$ and $\boldsymbol{z}_{(i)}$, for each $i = 1, \dots, m$. Sets $\hat{U}$ and U^i are disjoint for each $i = 1, \dots, m$ so this is equivalent to

$$\begin{aligned} &\max_{S \in C_k^-} \Big\{ \sum_{p \in S} \max_{\boldsymbol{z}_{(p)} \in U^p} \big(\alpha(S, p) \varphi_p(\hat{\boldsymbol{z}}, \boldsymbol{z}_{(p)}) - \sum_{l=k+1}^{n_y} \beta(S, l) y_l(\boldsymbol{z}) \big) \Big\} \le y_k(\boldsymbol{z}) \\ &\qquad \le \min_{T \in C_k^+} \Big\{ \sum_{q \in T} \min_{\boldsymbol{z}_{(q)} \in U^q} \big(\alpha(T, q) \varphi_q(\hat{\boldsymbol{z}}, \boldsymbol{z}_{(q)}) - \sum_{l=k+1}^{n_y} \beta(T, l) y_l(\boldsymbol{z}) \big) \Big\}, \quad \forall \hat{\boldsymbol{z}} \in \hat{U}. \end{aligned} \tag{A.6}$$

We proceed by backward induction. For $k = n_y$, i.e., the last eliminated variable, bounds (A.6) depend only on $\boldsymbol{z}$ and not on other adaptive variables. According to Lemma 10, each term $\varphi_i(\boldsymbol{z}_{(i)})$, $i = 1, \dots, m$, appears in upper bounds with a positive coefficient and in lower bounds with a negative coefficient for all variables $y_1(\boldsymbol{z}), \dots, y_{n_y}(\boldsymbol{z})$ (if it appears), or vice versa. Hence, the worst-case scenario for $\boldsymbol{z}_{(i)} \in U^i$ (in terms of feasibility) is equal for all linear terms in the lower and the upper bound for all $i = 1, \dots, m$. Plugging in this worst-case scenario yields lower and upper bounds on $y_{n_y}(\boldsymbol{z})$ depending only on $\hat{\boldsymbol{z}}$. Thus, there exists a decision rule for $y_{n_y}(\cdot)$ that is a function of only the non-constraintwise uncertain parameters $\hat{\boldsymbol{z}}$.

Suppose that for some k the lower and upper bounds (A.6) for $y_k(\boldsymbol{z})$ depend only on $\hat{\boldsymbol{z}}$. Thus, there exists a decision rule for $y_k(\cdot)$ that is a function of only $\hat{\boldsymbol{z}}$. Plug this decision rule in the lower and upper bounds (A.6) for $y_{k-1}(\boldsymbol{z})$. Then, according to Lemma 10, each term $\varphi_i(\boldsymbol{z}_{(i)})$, $i = 1, \dots, m$, appears in upper bounds with a positive coefficient and in lower bounds with a negative coefficient (if it appears), or vice versa. Hence, the worst-case scenario for $\boldsymbol{z}_{(i)} \in U^i$ (in terms of feasibility) is equal for all linear terms in the lower and the upper bound, for all $i = 1, \dots, m$. Plugging in this worst-case scenario yields lower and upper bounds on $y_{k-1}(\boldsymbol{z})$ depending only on $\hat{\boldsymbol{z}}$. This completes the induction.

Let $\boldsymbol{y}(\hat{\boldsymbol{z}})$ be the decision rule resulting from the above procedure. Because $\boldsymbol{x}$ is ARF to P_{hybrid}, the resulting pair $(\boldsymbol{x}, \boldsymbol{y}(\hat{\boldsymbol{z}}))$ is ARF to P_{hybrid}.

A.3 Proof Corollary 1

We note that if (3) has hybrid uncertainty and the objective (3a) contains adaptive variables, it can equivalently be written as

$$\min_{t,\boldsymbol{x},\boldsymbol{y}(\cdot)} \quad t, \tag{A.7a}$$

$$\text{s.t.} \quad \boldsymbol{c}(\hat{\boldsymbol{z}}, \boldsymbol{z}_{(0)})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{z}) \leq t \quad \forall (\hat{\boldsymbol{z}}, \boldsymbol{z}_{(0)}) \in \hat{U} \times U^0, \tag{A.7b}$$

$$\boldsymbol{a}_i(\hat{\boldsymbol{z}}, \boldsymbol{z}_{(i)})^\top \boldsymbol{x} + \boldsymbol{b}_i^\top \boldsymbol{y}(\boldsymbol{z}) \leq r_i(\hat{\boldsymbol{z}}, \boldsymbol{z}_{(i)}), \quad \forall (\hat{\boldsymbol{z}}, \boldsymbol{z}_{(i)}) \in \hat{U} \times U^i, \quad \forall i = 1, \dots, m, \tag{A.7c}$$

where $t \in \mathbb{R}$ is an auxiliary here-and-now decision variable. Problem (A.7) also has hybrid uncertainty, and a pair $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is ARO to (3) if and only if there exists a $t \in \mathbb{R}$ such that $(\boldsymbol{x}, \boldsymbol{y}(\cdot), t)$ is ARO to (A.7). Thus, in the remainder of the proof we can assume $\boldsymbol{d} = \boldsymbol{0}$, i.e., the objective is independent of adaptive variables.

According to Lemma 1, for any ARF $\boldsymbol{x}$ there exists a decision rule $\boldsymbol{y}(\cdot)$ that depends only on $\hat{\boldsymbol{z}}$ such that $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ is ARF to P_{hybrid}. Any $\boldsymbol{x}^*$ that is ARO to P_{hybrid} is also ARF to P_{hybrid}, so also for each ARO $\boldsymbol{x}^*$ there exists such a decision rule $\boldsymbol{y}^*(\cdot)$. The objective is independent of adaptive variables, so $(\boldsymbol{x}^*, \boldsymbol{y}(\cdot))$ is ARO for any ARF $\boldsymbol{y}(\cdot)$. Hence, $(\boldsymbol{x}^*, \boldsymbol{y}^*(\cdot))$ is ARO to P_{hybrid}.

A.4 Proof Lemma 2

We consider only adaptive robust feasibility and not optimality, so the objective of P_{block} can be ignored. Remove index 0 from its constraint set $K(v)$ (for some v). The set of constraints can be written as

$$\boldsymbol{a}_i(\boldsymbol{z}_{(v)})^\top \boldsymbol{x} + \boldsymbol{b}_i^\top \boldsymbol{y}_{(v)}(\boldsymbol{z}) \leq r_i(\boldsymbol{z}_{(v)}), \quad \forall \boldsymbol{z} \in U, \ \forall i \in K(v), \ \forall v = 1, \dots, V.$$

Due to the block uncertainty structure, all adaptive variables can be eliminated by performing FME on each block v separately. According to Lemma 10, bounds on each adaptive variable $y_k(\boldsymbol{z})$ can be represented by (A.1). If for some $k = 1, \dots, n_y$, variable $y_k(\boldsymbol{z})$ is an element of $\boldsymbol{y}_{(v)}(\boldsymbol{z})$ for some block v, any $S \in C_k^-$ or $T \in C_k^+$ is a subset of $K_{(v)}$, the original set of constraints for block v. The following two observations immediately follow for the given block v:

- For each $l = 1, \dots, n_y$ the coefficient of $y_l(\boldsymbol{z})$ is zero if $y_l(\boldsymbol{z})$ is not an element of $\boldsymbol{y}_{(v)}$, i.e., $\beta(S, l) = 0$ for all $S \in C_k^- \cup C_k^+$.
- For any p in S or T it holds that $\varphi_p(\cdot)$ is a function of $\boldsymbol{z}_{(v)}$ only.

For $k = n_y$, i.e., the last eliminated variable, this implies the lower and upper bounds on $y_{n_y}()$ are independent of $\boldsymbol{z}_{(w)}$ for $w \neq v$, and any feasible decision rule can be written as a function of $\boldsymbol{z}_{(v)}$ only. Plugging any such decision rule in the lower and upper bounds for $k = n_y - 1$ yields the same result for $y_{n_y-1}()$. The final result follows from backward induction.

Let $\boldsymbol{y}(\boldsymbol{z})$ be the decision rule resulting from the above procedure. Because $\boldsymbol{x}$ is ARF to P_{block}, the resulting pair $(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{z}))$ is ARF to P_{block}.

A.5 Proof Lemma 3

We consider only adaptive robust feasibility and not optimality, so the objective of P_{simplex} can be ignored. According to Lemma 10, in the FME procedure the bounds on variable $y_k(\boldsymbol{z})$ are given by (A.1). It is sufficient to satisfy the bounds on $y_k(\boldsymbol{z})$ for all extreme points of uncertainty set U, so we can alternatively write:

$$\max_{S_k \in C_k^-} \Big\{ \sum_{p \in S_k} \alpha(S_k, p) \varphi_p(\boldsymbol{x}, \boldsymbol{z}^j) - \sum_{l=k+1}^{n_y} \beta(S_k, l) y_l(\boldsymbol{z}^j) \Big\} \leq y_k(\boldsymbol{z}^j)$$
$$\leq \min_{T_k \in C_k^+} \Big\{ \sum_{q \in T} \alpha(T_k, q) \varphi_q(\boldsymbol{x}, \boldsymbol{z}^j) - \sum_{l=k+1}^{n_y} \beta(T_k, l) y_l(\boldsymbol{z}^j) \Big\}, \quad \forall \boldsymbol{z}^j, \ j = 1, \dots, L+1. \tag{A.8}$$

For each $j = 1, \dots, L+1$, let $l_k(\boldsymbol{z}^j)$ and $u_k(\boldsymbol{z}^j)$ denote the lower resp. upper bound on $y_k(\boldsymbol{z}^j)$ from (A.8). Affine independence of $\boldsymbol{z}^1, \dots, \boldsymbol{z}^{L+1}$ implies linear independence of $(1, \boldsymbol{z}^1), \dots, (1, \boldsymbol{z}^{L+1})$. Hence, by basic linear algebra, there exists exactly one $(a_0, \boldsymbol{a}) \in \mathbb{R} \times \mathbb{R}^L$ such that $a_0 + \boldsymbol{a}^\top \boldsymbol{z}^j = l(\boldsymbol{z}^j)$ for all $j = 1, \dots, L+1$.

Consider the LDR $y_k(\boldsymbol{z}) = a_0 + \boldsymbol{a}^\top \boldsymbol{z}$. Then $l(\boldsymbol{z}^j) = y_k(\boldsymbol{z}^j) \leq u(\boldsymbol{z}^j)$ for all $j = 1, \dots, L+1$. Hence, $y_k(\boldsymbol{z})$ is an LDR that satisfies bounds (A.8). Alternatively, one can construct an LDR that passes through points $(\boldsymbol{z}^j, u(\boldsymbol{z}^j))$ for all $j = 1, \dots, L+1$, or any LDR that is a convex combination of the previous two LDRs.

Thus, we can construct a decision rule for $y_k(\boldsymbol{z})$ that is linear in $\boldsymbol{z}$. For all $k = 1, \dots, n_y - 1$, this decision rule depends on $y_{k+1}(\boldsymbol{z}), \dots, y_{n_y}(\boldsymbol{z})$. For variable $y_{n_y}(\cdot)$, the constructed decision rule is independent of other adaptive variables. Plugging this in the decision rule for $y_{n_y-1}(\cdot)$ yields a decision rule that is again independent of other adaptive variables, and still linear in $\boldsymbol{z}$ because the coefficient for $y_{n_y}(\boldsymbol{z})$ in $l_{n_y-1}(\boldsymbol{z})$ and $u_{n_y-1}(\boldsymbol{z})$ does not depend on $\boldsymbol{z}$ (fixed recourse). Continuing this procedure yields LDRs for all adaptive variables $y_1(\cdot), \dots, y_{n_y}(\cdot)$.

Let $\boldsymbol{y}(\boldsymbol{z})$ be the decision rule resulting from the above procedure. Because $\boldsymbol{x}$ is ARF to P_{simplex}, the resulting pair $(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{z}))$ is ARF to P_{simplex}.

A.6 Proof Lemma 4

We first prove that the original problem (3) is equivalent to a convex PWL static RO problem; its proof uses Lemma 10.

Lemma 11. *If $(\boldsymbol{x}^*, \boldsymbol{y}^*(\cdot))$ is ARO to* (3), *$\boldsymbol{y}^*(\cdot)$ satisfies*

$$\boldsymbol{d}^\top \boldsymbol{y}^*(\boldsymbol{z}) = \max_{(S,T)\in M} \{h_{S,T}(\boldsymbol{x}^*, \boldsymbol{z})\}, \quad \forall \boldsymbol{z} \in U, \tag{A.9}$$

and $\boldsymbol{x}^*$ is optimal to

$$\min_{\boldsymbol{x}\in\mathcal{X}_{FME}} \quad \max_{\boldsymbol{z}\in U} \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \max_{(S,T)\in M} \{h_{S,T}(\boldsymbol{x}, \boldsymbol{z})\}, \tag{A.10}$$

with

$$M = \{(S,T) \mid \exists k = 1, \dots, n_y \ s.t. \ S \in C_k^-, T \in C_k^+, \beta(S,l) = \beta(T,l), \ \forall l > k, \ 0 \in S \cup T\},$$

and linear functions

$$h_{S,T}(\boldsymbol{x}, \boldsymbol{z}) = \sum_{p\in S, p>0} \frac{\alpha(S,p)}{\alpha(T,0) - \alpha(S,0)} \varphi_p(\boldsymbol{x}, \boldsymbol{z}) - \sum_{q\in T, q>0} \frac{\alpha(T,q)}{\alpha(T,0) - \alpha(S,0)} \varphi_q(\boldsymbol{x}, \boldsymbol{z}),$$

and sets C^-, C^+, functions $\varphi(\cdot)$ and coefficients α and β defined as in Lemma 10. Conversely, if $\boldsymbol{x}^*$ is optimal to (A.10)*, there exists a $\boldsymbol{y}^*(\cdot)$ such that $(\boldsymbol{x}^*, \boldsymbol{y}^*(\cdot))$ is ARO to* (3)*, and any such $\boldsymbol{y}^*(\cdot)$ satisfies* (A.9)*.*

Proof of Lemma 11. Consider problem (3), with the objective moved to the constraints using epigraph variable $t \in \mathbb{R}$:

$$\begin{aligned} \min_{t,\boldsymbol{x},\boldsymbol{y}(\cdot)} \quad & t, && \text{(A.11a)} \\ \text{s.t.} \quad & t \geq \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U, && \text{(A.11b)} \\ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \leq \boldsymbol{r}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U. && \text{(A.11c)} \end{aligned}$$

Eliminate all adaptive variables in (A.11b)-(A.11c) via FME. Let $\varphi_0(\boldsymbol{x}, t, \boldsymbol{z}) = t - \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x}$. In notation of Lemma 10, FME is performed on

$$\begin{aligned} & \boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{z}) \leq \varphi_0(\boldsymbol{x}, t, \boldsymbol{z}), && \text{(A.12a)} \\ & \boldsymbol{b}_i^\top \boldsymbol{y}(\boldsymbol{z}) \leq \varphi_i(\boldsymbol{x}, t, \boldsymbol{z}), \ \forall i = 1, \dots, m, && \text{(A.12b)} \end{aligned}$$

where the coefficient for t is zero in φ_i, $i = 1, \dots, m$. According to Lemma 10, after elimination of variable k, inequalities (A.1) hold. Suppose for some $S_k \in C_k^-$, $T_k \in C_k^+$ the upper and lower bounds on $y_k(\boldsymbol{z})$ do not depend on $y_{k+1}(\boldsymbol{z}), \dots, y_{n_y}(\boldsymbol{z})$. Then the following constraint is derived for the static robust optimization problem after completing the full FME procedure:

$$\sum_{p\in S_k} \alpha(S_k, p)\varphi_p(\boldsymbol{x}, t, \boldsymbol{z}) \leq \sum_{q\in T_k} \alpha(T_k, q)\varphi_q(\boldsymbol{x}, t, \boldsymbol{z}), \quad \forall \boldsymbol{z} \in U, \tag{A.13}$$

where $\varphi_p(\cdot)$ is a function of t only if $p = 0$. Constraints of the original system (A.11c) that are independent of adaptive variables can also be represented in form (A.13). Original constraints (A.11b) are part of a particular constraint in form (A.13) if and only if $0 \in S_k \cup T_k$ for some $S_k \in C_k^-$, $T_k \in C_k^+$, $k = 1, \ldots, n_y$. Thus, problem (A.11) after FME can be written as

$$\min_{t,\boldsymbol{x}} \quad t, \tag{A.14a}$$

$$\text{s.t.} \quad \sum_{p\in S} \alpha(S,p)\varphi_p(\boldsymbol{x},t,\boldsymbol{z}) \le \sum_{q\in T} \alpha(T,q)\varphi_q(\boldsymbol{x},t,\boldsymbol{z}), \quad \forall (S,T)\in M, \quad \forall \boldsymbol{z}\in U, \tag{A.14b}$$

$$\sum_{p\in S} \alpha(S,p)\varphi_p(\boldsymbol{x},t,\boldsymbol{z}) \le \sum_{q\in T} \alpha(T,q)\varphi_q(\boldsymbol{x},t,\boldsymbol{z}), \quad \forall (S,T)\in N, \quad \forall \boldsymbol{z}\in U, \tag{A.14c}$$

with

$$M = \{(S,T) \mid \exists k = 1,\ldots,n_y \text{ s.t. } S\in C_k^-, T\in C_k^+, \beta(S,l) = \beta(T,l), \ \forall l > k, \ 0 \in S\cup T\}, \tag{A.15a}$$

$$N = \{(S,T) \mid \exists k = 1,\ldots,n_y \text{ s.t. } S\in C_k^-, T\in C_k^+, \beta(S,l) = \beta(T,l), \ \forall l > k, \ 0 \notin S\cup T\}. \tag{A.15b}$$

In other words, we separated the constraints depending on t from the constraints not depending on t. From Lemma 10 one can see that (A.14c) is the result of performing FME on the set of constraints (A.12b), which are the constraints defining set $\mathcal{X}$. Thus, (A.14c) describes set $\mathcal{X}_{\text{FME}}$. Furthermore, if we define $\alpha(S,0) = 0$ if $0 \notin S$ and $\gamma(T,0) = 0$ if $0 \notin T$, constraint (A.14b) can be rewritten to

$$\begin{aligned} t \ge \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \sum_{p\in S, p>0} \frac{\alpha(S,p)}{\alpha(T,0)-\alpha(S,0)}\varphi_p(\boldsymbol{x},t,\boldsymbol{z}) - \sum_{q\in T, q>0} \frac{\alpha(T,q)}{\alpha(T,0)-\alpha(S,0)}\varphi_q(\boldsymbol{x},t,\boldsymbol{z}) \\ \forall (S,T)\in M, \quad \forall \boldsymbol{z}\in U, \end{aligned} \tag{A.16}$$

because $\alpha(T,0) > \alpha(S,0)$ according to Lemma 10. Note that the coefficient for t is zero for all functions φ on the RHS. Thus, for fixed $\boldsymbol{z}\in U$, constraint (A.16) defines a lower bound on epigraph variable t that is convex PWL in $\boldsymbol{x}$. Subsequently, we eliminate t and define

$$h_{S,T}(\boldsymbol{x},\boldsymbol{z}) = \sum_{p\in S, p>0} \frac{\alpha(S,p)}{\alpha(T,0)-\alpha(S,0)}\varphi_p(\boldsymbol{x},\boldsymbol{z}) - \sum_{q\in T, q>0} \frac{\alpha(T,q)}{\alpha(T,0)-\alpha(S,0)}\varphi_q(\boldsymbol{x},\boldsymbol{z}). \tag{A.17}$$

This yields the following problem equivalent to (A.14):

$$\min_{\boldsymbol{x}\in\mathcal{X}_{\text{FME}}} \ \max_{\boldsymbol{z}\in U} \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x} + \max_{(S,T)\in M} \{h_{S,T}(\boldsymbol{x},\boldsymbol{z})\}. \tag{A.18}$$

If $(\boldsymbol{x}^*, t^*, \boldsymbol{y}^*(\cdot))$ is optimal to (A.11), $\boldsymbol{x}^*$ is optimal to (A.18) with equal objective value. This implies that $\boldsymbol{y}^*(\cdot)$ satisfies

$$\boldsymbol{d}^\top \boldsymbol{y}^*(\boldsymbol{z}) = \max_{(S,T)\in M} \{h_{S,T}(\boldsymbol{x}^*,\boldsymbol{z})\}, \quad \forall \boldsymbol{z}\in U. \tag{A.19}$$

Conversely, if $\boldsymbol{x}^*$ is optimal to (A.18), there exists a $(t^*, \boldsymbol{y}^*(\cdot))$ such that $(\boldsymbol{x}^*, t^*, \boldsymbol{y}^*(\cdot))$ is optimal to (A.11) with equal objective value. This implies that any such $\boldsymbol{y}^*(\cdot)$ satisfies (A.19). Lastly, note that $\boldsymbol{x}^*$ is optimal to (3) if and only if there exists a $t^* \in \mathbb{R}$ such that $(t^*, \boldsymbol{x}^*)$ is optimal to (A.11). This completes the proof. □

The result of Lemma 11 is also illustrated in Example 2, where if auxiliary variable t is eliminated the resulting problem has a convex PWL objective. If the number of adaptive variables in (3) is small enough that full FME can be performed (order of magnitude: 20 adaptive variables (Zhen et al., 2018a)), one can solve (10) via an epigraph formulation in order to obtain an ARO $\boldsymbol{x}$ to (3).

We are now in position to prove the result of Lemma 4.

Proof of Lemma 4. By Definition 7(i) a solution $\boldsymbol{x}^*$ is PARO to (3) if and only if

- There exists a $\boldsymbol{y}^* \in \mathcal{R}^{L,n_y}$ such that $(\boldsymbol{x}^*, \boldsymbol{y}^*(\cdot))$ is ARO to (3) and there does not exist a pair $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\cdot))$ that is ARO to (3) and the following conditions hold:

$$\begin{aligned} \boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}(\boldsymbol{z}) &\le \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x}^* + \boldsymbol{d}^\top \boldsymbol{y}^*(\boldsymbol{z}), \quad \forall \boldsymbol{z}\in U, \\ \boldsymbol{c}(\bar{\boldsymbol{z}})^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}(\bar{\boldsymbol{z}}) &< \boldsymbol{c}(\bar{\boldsymbol{z}})^\top \boldsymbol{x}^* + \boldsymbol{d}^\top \boldsymbol{y}^*(\bar{\boldsymbol{z}}), \quad \text{for some } \bar{\boldsymbol{z}}\in U. \end{aligned} \tag{A.20}$$

By Lemma 11, this holds if and only if

- $\boldsymbol{x}^*$ is optimal to (10) and there exists a $\boldsymbol{y}^* \in \mathcal{R}^{L,n_y}$ such that

$$\boldsymbol{d}^\top \boldsymbol{y}^*(\boldsymbol{z}) = \max_{(S,T)\in M} \{h_{S,T}(\boldsymbol{x}^*, \boldsymbol{z})\} \quad \forall \boldsymbol{z} \in U, \tag{A.21}$$

  and there does not exist a $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ such that $\bar{\boldsymbol{x}}$ is optimal to (10) and $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\cdot))$ satisfies (A.21) and (A.20) holds.

Substituting (A.21) in (A.20) yields the following set of equivalent conditions:

- $\boldsymbol{x}^*$ is optimal to (10) and there does not exist another $\bar{\boldsymbol{x}}$ optimal to (10) such that

$$\boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}} + \max_{(S,T)\in M} \{h_{S,T}(\bar{\boldsymbol{x}}, \boldsymbol{z})\} \leq \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x}^* + \max_{(S,T)\in M} \{h_{S,T}(\boldsymbol{x}^*, \boldsymbol{z})\}, \quad \forall \boldsymbol{z} \in U,$$
$$\boldsymbol{c}(\bar{\boldsymbol{z}})^\top \bar{\boldsymbol{x}} + \max_{(S,T)\in M} \{h_{S,T}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{z}})\} < \boldsymbol{c}(\bar{\boldsymbol{z}})^\top \boldsymbol{x}^* + \max_{(S,T)\in M} \{h_{S,T}(\boldsymbol{x}^*, \bar{\boldsymbol{z}})\}, \quad \text{for some } \bar{\boldsymbol{z}} \in U.$$

This statement holds if and only if $\boldsymbol{x}^*$ is PRO to (10), by Definition 1. □

A.7 Proof Theorem 1

First, we prove the existence of PRO solutions to a general class of static RO problems, with bounded feasible region $\mathcal{X}$. This boundedness assumption cannot be relaxed. For example, consider the static RO problem $\max_{x\geq 0} \min_{z\in[0,1]} xz$. The worst-case scenario is $z = 0$, and any $x \geq 0$ is worst-case optimal. In any other scenario $z > 0$, higher x is better. Any x is dominated by $x + \epsilon$ with $\epsilon > 0$, and there is no PRO solution.

Lemma 12. *Let $f : \mathbb{R}^n \times \mathbb{R}^L \mapsto \mathbb{R}$, with $f(\boldsymbol{x}, \boldsymbol{z})$ continuous in $\boldsymbol{z}$. Consider the static RO problem*

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{z}\in U} \; f(\boldsymbol{x}, \boldsymbol{z}). \tag{A.22}$$

Let $U \subseteq \mathbb{R}^L$ be closed, convex with a nonempty relative interior. If (i) $\mathcal{X}$ is compact and $f(\boldsymbol{x}, \boldsymbol{z})$ continuous in $\boldsymbol{x}$ and/or (ii) $\mathcal{X}$ is a finite set, and additionally there exists an RO solution to (A.22)*, there also exists a PRO solution to* (A.22).

Proof of Lemma 12. Let $(\mathbb{R}^L, \mathcal{B}(\mathbb{R}^L))$ be a measurable space, with $\mathcal{B}(\mathbb{R}^L)$ the Borel σ-algebra. For fixed $\boldsymbol{x}$, function $f(\boldsymbol{x}, \boldsymbol{z})$ is continuous in $\boldsymbol{z}$, so it is measurable on closed subsets of $\mathbb{R}^L$, in particular set U. Define function $g : \mathbb{R}^n \mapsto \mathbb{R}$ with

$$g(\boldsymbol{x}) := \int_U f(\boldsymbol{x}, \boldsymbol{z}) dP(\boldsymbol{z}), \tag{A.23}$$

where P denotes a strictly positive probability measure on $\mathbb{R}^L$, such as the Gaussian measure. Because $0 \leq P(U) \leq P(\mathbb{R}^L) = 1$, the Lebesgue integral (A.23) assumes finite values for any $\boldsymbol{x}$. Hence, $f(\boldsymbol{x}, \boldsymbol{z})$ is Lebesgue-integrable in its second argument on measured space $(\mathbb{R}^L, \mathcal{B}(\mathbb{R}^L), P)$ for any $\boldsymbol{x}$ and g is well-defined.

We proceed by showing that an optimal solution to the following optimization problem is PRO to (A.22):

$$\min_{\boldsymbol{x}\in\mathcal{X}^{\mathrm{RO}}} g(\boldsymbol{x}). \tag{A.24}$$

The remainder of the proof consists of two parts. First, we show that an optimal solution to (A.24) is always attained. Subsequently, we show that such an optimal solution is PRO to (A.22).

Part 1 (The optimum is attained):
We treat the two cases for $\mathcal{X}$ separately.

Case (i): Set $\mathcal{X}$ is compact and $f(\boldsymbol{x}, \boldsymbol{z})$ continuous in $\boldsymbol{x}$. We show that g is continuous. Consider a sequence $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ converging to $\boldsymbol{x}$. By continuity of f in $\boldsymbol{x}$, $\lim_{n\to\infty} f(\boldsymbol{x}_n, \boldsymbol{z}) = f(\boldsymbol{x}, \boldsymbol{z})$. Thus,

$$g(\boldsymbol{x}) = \int_U f(\boldsymbol{x}, \boldsymbol{z}) dP(\boldsymbol{z}) = \int_U \lim_{n\to\infty} f(\boldsymbol{x}_n, \boldsymbol{z}) dP(\boldsymbol{z}). \tag{A.25}$$

Let $M > 0$ be such that $|f(\boldsymbol{x},\boldsymbol{z})| < M$, and define $h : \mathbb{R}^L \mapsto \mathbb{R}$ with $h(\boldsymbol{z}) = M$ for all $\boldsymbol{z}$. Then h is Lebesgue-integrable, and we can apply the dominated convergence theorem to switch the order of the limit and integration in (A.25) to obtain

$$g(\boldsymbol{x}) = \lim_{n\to\infty} \int_U f(\boldsymbol{x}_n,\boldsymbol{z})dP(\boldsymbol{z}) = \lim_{n\to\infty} g(\boldsymbol{x}_n),$$

Hence, $g(\boldsymbol{x})$ is continuous for each $\boldsymbol{x} \in \mathbb{R}^n$. Let $\mathcal{X}^{\text{RO}}$ denote the set of robustly (worst-case) optimal solutions to (A.22). Then $\mathcal{X}^{\text{RO}}$ is compact if $\mathcal{X}$ is compact. Problem (A.24) minimizes a continuous function over a compact domain, so, by the extreme value theorem, a minimum is always attained.

Case (ii): Set $\mathcal{X}$ is a finite set. Problem (A.24) minimizes $g(\boldsymbol{x})$ over a finite set, so the minimum is attained.

Part 2 (An optimal solution is PRO):
Let $\hat{\boldsymbol{x}}$ denote an optimal solution to (A.24). We proceed by showing via proof by contradiction that $\hat{\boldsymbol{x}}$ is PRO to (A.22). Suppose $\hat{\boldsymbol{x}}$ is not PRO to (A.22). Then there exists an $\bar{\boldsymbol{x}} \in \mathcal{X}^{\text{RO}}$ such that

$$\begin{aligned} f(\bar{\boldsymbol{x}},\boldsymbol{z}) &\le f(\hat{\boldsymbol{x}},\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U,\\ f(\bar{\boldsymbol{x}},\bar{\boldsymbol{z}}) &< f(\hat{\boldsymbol{x}},\bar{\boldsymbol{z}}), \quad \text{for some } \bar{\boldsymbol{z}} \in U. \end{aligned}$$

We proceed by showing that there must exist a ball contained in U with strictly positive measure where strict inequality holds. Let $\bar{B}$ denote the ball with radius δ centered at $\bar{\boldsymbol{z}}$:

$$\bar{B} = \{\boldsymbol{z} \in \mathbb{R}^L : \|\boldsymbol{z} - \bar{\boldsymbol{z}}\|_2 \le \delta\}.$$

By continuity of $f(\bar{\boldsymbol{x}},\boldsymbol{z}) - f(\hat{\boldsymbol{x}},\boldsymbol{z})$ w.r.t. $\boldsymbol{z}$, there exists a $\delta > 0$ such that for each $\boldsymbol{z} \in \bar{B}$ it holds that $f(\bar{\boldsymbol{x}},\boldsymbol{z}) - f(\hat{\boldsymbol{x}},\boldsymbol{z}) < 0$. Note that $\bar{\boldsymbol{z}}$ need not be in the relative interior of U. Hence, the ball $\bar{B}$ need not be contained in U. Let $\tilde{\boldsymbol{z}} \in \text{ri}(U)$. We construct a new scenario $\boldsymbol{z}^* = \theta\tilde{\boldsymbol{z}} + (1-\theta)\bar{\boldsymbol{z}}$. Because U is convex, $\boldsymbol{z}^* \in \text{ri}(U)$ if $0 \le \theta < 1$ according to Rockafellar (1970, Theorem 6.1). Choosing $1 - \delta\|\tilde{\boldsymbol{z}} - \bar{\boldsymbol{z}}\|_2^{-1} < \theta < 1$ ensures that $\boldsymbol{z}^* \in \text{int}(\bar{B}) \cap \text{ri}(U) = \text{ri}(U \cap \bar{B})$. Consider the ball B^* with radius $\epsilon > 0$ centered at $\boldsymbol{z}^*$:

$$B^* = \{\boldsymbol{z} \in \mathbb{R}^L : \|\boldsymbol{z} - \boldsymbol{z}^*\|_2 \le \epsilon\}.$$

For sufficiently small $\epsilon > 0$, it holds that $\boldsymbol{z} \in B^* \Rightarrow \boldsymbol{z} \in U \cap \bar{B}$. In other words, for such an ϵ, each point $\boldsymbol{z} \in B^*$ is in the uncertainty set U and is such that $f(\bar{\boldsymbol{x}},\boldsymbol{z}) < f(\hat{\boldsymbol{x}},\boldsymbol{z})$.

Finally, we consider the difference between $g(\bar{\boldsymbol{x}})$ and $g(\hat{\boldsymbol{x}})$ on U. Note that $|g(\boldsymbol{x})| < \infty$ for all $\boldsymbol{x}$. The following holds:

$$g(\bar{\boldsymbol{x}}) - g(\hat{\boldsymbol{x}}) = \int_{U\setminus B^*} f(\bar{\boldsymbol{x}},\boldsymbol{z}) - f(\hat{\boldsymbol{x}},\boldsymbol{z})dP(\boldsymbol{z}) + \int_{B^*} f(\bar{\boldsymbol{x}},\boldsymbol{z}) - f(\hat{\boldsymbol{x}},\boldsymbol{z})dP(\boldsymbol{z}).$$

The first integral is nonpositive since $f(\bar{\boldsymbol{x}},\boldsymbol{z}) \le f(\hat{\boldsymbol{z}},\boldsymbol{z})$ for each $\boldsymbol{z} \in U\setminus B^*$. The second integral is strictly negative since $f(\bar{\boldsymbol{x}},\boldsymbol{z}) < f(\hat{\boldsymbol{z}},\boldsymbol{z})$ for $\boldsymbol{z} \in B^*$ and measure P is strictly positive, i.e., $P(B^*) > 0$. Hence, $g(\bar{\boldsymbol{x}}) < g(\hat{\boldsymbol{x}})$, contradicting the fact that $\hat{\boldsymbol{x}}$ is optimal to (A.24). □

The result of Theorem 1 immediately follows.

Proof of Theorem 1. By Lemma 4, it suffices to prove existence of a PRO solution to (10). Because $\mathcal{X} = \mathcal{X}_{\text{FME}}$, set $\mathcal{X}_{\text{FME}}$ is compact. By construction of (3), uncertainty set U is assumed to be convex, compact with a nonempty relative interior. Lastly, the objective function of (10) is continuous in $\boldsymbol{x}$ and $\boldsymbol{z}$. Hence, all conditions of Lemma 12 are satisfied, and existence of a PARO solution to (3) is guaranteed. □

A.8 Proof Lemma 5 via FME

Let $\boldsymbol{x}$ be ARF to (3). W.l.o.g., suppose in the FME procedure the adaptive variables are eliminated in the order $y_1,\ldots,y_{n_y}$, i.e., according to their index. Let $F_k(y_{k+1}(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z}),\boldsymbol{z})$ denote the optimal decision rule for y_k as a function of the decision rules for the adaptive variables with higher index and the uncertain parameter $\boldsymbol{z}$. We prove by induction on $k = 1,\ldots,n_y$ that $F_k(y_{k+1}(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z}),\boldsymbol{z})$ is jointly PWL in $y_{k+1},\ldots,y_{n_y}$ and $\boldsymbol{z}$.

According to Lemma 10, we can write the bounds after elimination of variable $y_1(\boldsymbol{z})$ as

$$\max_{S\in C_1^-}\Big\{\sum_{p\in S}\alpha(S,p)\varphi_p(\boldsymbol{z})-\sum_{l=2}^{n_y}\beta(S,l)y_l(\boldsymbol{z})\Big\}\le y_1(\boldsymbol{z})$$
$$\le\min_{T\in C_1^+}\Big\{\sum_{q\in T}\alpha(T,q)\varphi_q(\boldsymbol{z})-\sum_{l=2}^{n_y}\beta(T,l)y_l(\boldsymbol{z})\Big\},\quad\forall\boldsymbol{z}\in U,$$

for some coefficients α and β independent of $\boldsymbol{z}$. For fixed $y_2,\ldots,y_{n_y}$, $\boldsymbol{z}$ and $\boldsymbol{x}$, the highest possible contribution of y_1 to the objective value is achieved by setting y_1 equal to its upper bound if $d_1<0$, and equal to its lower bound if $d_1>0$. Thus, $F_1(y_2(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z}),\boldsymbol{z})$ is equal to either the upper or the lower bound on y_1. Both the upper and lower bound are jointly PWL in y_i, $i=2,\ldots,n_y$ and $\boldsymbol{z}$.

Now, suppose that for each $i=1,\ldots,k-1$, after elimination of variable $y_i(\boldsymbol{z})$ the optimal decision rule $F_i(y_{i+1}(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z}),\boldsymbol{z})$ is jointly PWL in $y_{i+1},\ldots,y_{n_y}$.

After elimination of $y_k(\boldsymbol{z})$ we can again write the bounds according to Lemma 10. For fixed $y_{k+1},\ldots,y_{n_y}$, $\boldsymbol{z}$ and $\boldsymbol{x}$, the highest possible contribution of y_k to the objective value is achieved by minimizing $\boldsymbol{d}^\top\boldsymbol{y}$, i.e., solving

$$\min_{y_k}\ \sum_{i=1}^{k-1}d_iF_i(F_{i+1}(\ldots),\ldots,F_{k-1}(y_k(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z}),\boldsymbol{z}),y_k(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z}),\boldsymbol{z})$$
$$+d_ky_k(\boldsymbol{z})+\sum_{i=k+1}^{n_y}d_iy_i(\boldsymbol{z}),\tag{A.26a}$$
$$\text{s.t.}\ \max_{S\in C_k^-}\Big\{\sum_{p\in S}\alpha(S,p)\varphi_p(\boldsymbol{z})-\sum_{l=k+1}^{n_y}\beta(S,l)y_l(\boldsymbol{z})\Big\}\le y_k(\boldsymbol{z}),\tag{A.26b}$$
$$\min_{T\in C_k^+}\Big\{\sum_{q\in T}\alpha(T,q)\varphi_q(\boldsymbol{z})-\sum_{l=k+1}^{n_y}\beta(T,l)y_l(\boldsymbol{z})\Big\}\ge y_k(\boldsymbol{z}),\tag{A.26c}$$

where the last term in the objective (the last summation) may be dropped because it does not depend on y_k. In the objective each decision rule F_i, $i=1,\ldots,k-1$, is a function of the decision rules $F_{i+1},\ldots,F_{k-1}$, variables $y_k(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z})$ and $\boldsymbol{z}$. Plugging in a PWL argument in a PWL function retains the piecewise linear structure. Thus, (A.26) asks to minimize a univariate PWL function on a closed interval. The optimum is attained at either an interior point or a boundary point; we consider these cases separately.

- Problem (A.26) has a boundary minimum. The minimum is attained at either the lower or upper bounds provided by (A.26b) and (A.26c). In this case, $F_k(y_{k+1}(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z}),\boldsymbol{z})$ is clearly jointly PWL in $y_{k+1}(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z})$ and $\boldsymbol{z}$.
- Problem (A.26) has an interior minimum. The unrestricted minimum of (A.26a) is at the intersection of two functions that are jointly linear in $y_k,\ldots,y_{n_y}$ and $\boldsymbol{z}$. Any intersection point can be expressed as

$$s_0(\boldsymbol{z})+\sum_{i=k}^{n_y}s_iy_i(\boldsymbol{z})=t_0(\boldsymbol{z})+\sum_{i=k}^{n_y}t_iy_i(\boldsymbol{z}),$$

  for some scalars $s_0(\boldsymbol{z})$ and $t_0(\boldsymbol{z})$ depending linearly on $\boldsymbol{z}$ and some vectors $\boldsymbol{s},\boldsymbol{t}\in\mathbb{R}^{n_y-k}$. This is equivalent to

$$y_k(\boldsymbol{z})=\frac{s_0(\boldsymbol{z})-t_0(\boldsymbol{z})+\sum_{i=k+1}^{n_y}(s_i-t_i)y_i(\boldsymbol{z})}{t_k-s_k},$$

  and this is jointly linear in $y_k,\ldots,y_{n_y}$ and $\boldsymbol{z}$. The pair $\{(s_0(\boldsymbol{z}),\boldsymbol{s}),(t_0(\boldsymbol{z}),\boldsymbol{t})\}$ that defines the interior minimum intersection point depends on $y_k,\ldots,y_{n_y}$ and $\boldsymbol{z}$. Thus, the optimal decision rule $F_k(y_{k+1}(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z}),\boldsymbol{z})$ is a PWL function of $y_{k+1},\ldots,y_{n_y}$ and $\boldsymbol{z}$.

This completes the induction step. Lastly, note that $F_{n_y}(\boldsymbol{z})$ is PWL in $\boldsymbol{z}$ and that plugging in a PWL argument in a PWL function retains the piecewise linear structure. Thus, going from $k=n_y$ to $k=1$ and for each k plugging in $F_k(y_{k+1}(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z}),\boldsymbol{z})$ in $F_{k-1}(y_k(\boldsymbol{z}),\ldots,y_{n_y}(\boldsymbol{z}),\boldsymbol{z})$ yields decision rules that are PWL in $\boldsymbol{z}$ for all variables $y_1,\ldots,y_{n_y}$.

A.9 Proof Lemma 5 via linear optimization

Let $\boldsymbol{x}$ be ARF to (3). We make use of the concept of basic solutions in linear optimization (Bertsimas and Tsitsiklis, 1997). In standard form the remaining problem for $\boldsymbol{y}$ for fixed $\boldsymbol{z}$, reads:

$$\min_{\boldsymbol{y}^+,\boldsymbol{y}^-,\boldsymbol{s}} \quad \boldsymbol{d}^\top(\boldsymbol{y}^+ - \boldsymbol{y}^-), \tag{A.27a}$$

$$\text{s.t.} \quad \boldsymbol{B}(\boldsymbol{y}^+ - \boldsymbol{y}^-) + \boldsymbol{s} = \boldsymbol{r}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}, \tag{A.27b}$$

$$\boldsymbol{y}^+, \boldsymbol{y}^-, \boldsymbol{s} \geq \boldsymbol{0}, \tag{A.27c}$$

where $\boldsymbol{s}$ is a slack variable and $\boldsymbol{y}$ is represented by the difference of two nonnegative variables. Let $\boldsymbol{v} \in \mathbb{R}^{2n_y+m}$, $\boldsymbol{M} \in \mathbb{R}^{m\times(2n_y+m)}$ and $\boldsymbol{f} \in \mathbb{R}^{2n_y+m}$ denote the vector of decision variables, the equality constraint matrix and the objective vector of (A.27), respectively:

$$\boldsymbol{v} = [\boldsymbol{y}^+ \ \boldsymbol{y}^- \ \boldsymbol{s}]^\top, \ \boldsymbol{M} = [\boldsymbol{B} \ \text{-}\boldsymbol{B} \ \boldsymbol{I}], \ \boldsymbol{f} = [\boldsymbol{d} \ \text{-}\boldsymbol{d} \ \boldsymbol{0}]^\top. \tag{A.28}$$

Each basis is represented by m linearly independent columns of $\boldsymbol{M}$. Let $\boldsymbol{W} \in \mathbb{R}^{m\times m}$ denote a basis matrix, and let $\boldsymbol{v}_{\boldsymbol{W}}$ and $\boldsymbol{f}_{\text{W}}$ denote the components of $\boldsymbol{v}$ and $\boldsymbol{f}$ corresponding to the basic variables. For any basic solution $\boldsymbol{v}$ it holds that

$$\boldsymbol{v}_{\boldsymbol{W}} = \boldsymbol{W}^{-1}(\boldsymbol{r}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}), \tag{A.29}$$

and the remaining non-basic components of $\boldsymbol{v}$ are equal to zero. Denote the basic solution by $(\boldsymbol{v}_{\boldsymbol{W}}, \boldsymbol{0}_{\setminus \boldsymbol{W}})$; it is a basic feasible solution (BFS) to (A.27) if and only if $\boldsymbol{v}_{\boldsymbol{W}} \geq \boldsymbol{0}$. For optimality of $(\boldsymbol{v}_{\boldsymbol{W}}, \boldsymbol{0}_{\setminus \boldsymbol{W}})$ it is additionally required that the reduced costs are nonnegative. Nonnegativity of the reduced costs (i.e., optimality of $(\boldsymbol{v}_{\boldsymbol{W}}, \boldsymbol{0}_{\setminus \boldsymbol{W}})$) reads

$$\boldsymbol{f} - \boldsymbol{f}_{\text{W}}^\top \boldsymbol{W}^{-1}\boldsymbol{M} \geq \boldsymbol{0}. \tag{A.30}$$

We restrict ourselves to those basic solutions for which optimality condition (A.30) holds, note that this condition is independent of $\boldsymbol{z}$. It follows that for each basis matrix $\boldsymbol{W}$ that satisfies (A.30), it associated basic solution $(\boldsymbol{v}_{\boldsymbol{W}}, \boldsymbol{0}_{\setminus \boldsymbol{W}})$ is feasible (and optimal) if and only if $\boldsymbol{z}$ is in the following subset of U:

$$U_{\boldsymbol{W}}(\boldsymbol{x}) = \{\boldsymbol{z} \in U \ : \ \boldsymbol{W}^{-1}(\boldsymbol{r}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}) \geq \boldsymbol{0}\}.$$

Let $\boldsymbol{y}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{W})$ denote the basic solution corresponding to $\boldsymbol{W}$ in terms of the original variables $\boldsymbol{y}$. From (A.29) it follows that $\boldsymbol{y}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{W})$ is linear in $\boldsymbol{z}$.

Any basic solution to (A.27) corresponds with at least one basis, and each basis is represented by m linearly independent columns of $\boldsymbol{M}$. Thus, there are at most $\beta = \binom{2n_y+m}{m}$ bases (i.e., matrices $\boldsymbol{W}$) to (A.27) that satisfy (A.30), independent of $\boldsymbol{z}$. Number the matrices $\boldsymbol{W}_1, \dots, \boldsymbol{W}_\beta$. Each of these matrices $\boldsymbol{W}_j$ has its own LDR $\boldsymbol{y}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{W}_j)$ that is optimal for all $\boldsymbol{z} \in U_{\boldsymbol{W}_j}(\boldsymbol{x})$.

Because $\boldsymbol{x}$ is ARF to (3) and (3) has a finite optimal objective value, problem (A.27) is feasible and has a finite optimum for all $\boldsymbol{z} \in U$. Therefore, there exists an optimal basic feasible solution for all $\boldsymbol{z} \in U$, and the union of all $U_{\boldsymbol{W}_i}$ equals U itself. This implies that, for the given $\boldsymbol{x}$, the following PWL decision rule is optimal for each $\boldsymbol{z} \in U$:

$$\boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{y}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{W}_{i^*}) \text{ if } i^* = \min\{i : \boldsymbol{z} \in U_{\boldsymbol{W}_i}(\boldsymbol{x})\}.$$

Note that a different numbering of the matrices gives a (possibly) different optimal PWL decision rule. In essence, the proof performs sensitivity analysis on the right-hand side vectors of (A.27), which is the only term in (A.27) that depends on $\boldsymbol{z}$.

A.10 Proof Theorem 3

Let OPT denote the optimal (worst-case) objective value of P. By Definition 7(i), and using that $\boldsymbol{d} = \boldsymbol{0}$, a solution $\boldsymbol{x}^*$ is PARO to P if and only if the following statement holds:

- There exists a $\boldsymbol{y}^* \in \mathcal{R}^{L,n_y}$ such that $(\boldsymbol{x}^*, \boldsymbol{y}^*(\cdot))$ is ARO to P and there does not exist a pair $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\cdot))$ that is ARO to P and

$$\begin{aligned} &\boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}} \leq \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x}^*, \quad \forall \boldsymbol{z} \in U, \\ &\boldsymbol{c}(\bar{\boldsymbol{z}})^\top \bar{\boldsymbol{x}} < \boldsymbol{c}(\bar{\boldsymbol{z}})^\top \boldsymbol{x}^*, \quad \text{for some } \bar{\boldsymbol{z}} \in U. \end{aligned} \tag{A.31}$$

By definition of set $\mathcal{X}$, this holds if and only if

- $\boldsymbol{x}^* \in \mathcal{X}$, OPT $= \max_{\boldsymbol{z}\in U} \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x}^*$ and there does not exist an $\bar{\boldsymbol{x}} \in \mathcal{X}$ such that OPT $= \max_{\boldsymbol{z}\in U} \boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}}$ and (A.31) holds.

Because for any ARF $\boldsymbol{x}$ there exists an ARF decision rule $\boldsymbol{y}(\cdot)$ such that $\boldsymbol{y}(\boldsymbol{z}) = f_{\boldsymbol{w}}(\boldsymbol{z})$ for some $\boldsymbol{w}$, it follows that $\mathcal{X}$ is equal to

$$\mathcal{X}_f = \{\boldsymbol{x} \in \mathbb{R}^{n_x} \mid \exists \boldsymbol{w} \in \mathbb{R}^p : \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B} f_{\boldsymbol{w}}(\boldsymbol{z}) \leq \boldsymbol{r}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U\},$$

which is the set of feasible $\boldsymbol{x}$ when Stage-2 decision rules are restricted to be of form $f_{\boldsymbol{w}}(\boldsymbol{z})$. Hence, the previous set of conditions holds if and only if

- $\boldsymbol{x}^* \in \mathcal{X}_f$, OPT $= \max_{\boldsymbol{z}\in U} \boldsymbol{c}(\boldsymbol{z})^\top \boldsymbol{x}^*$ and there does not exist an $\bar{\boldsymbol{x}} \in \mathcal{X}$ such that OPT $= \max_{\boldsymbol{z}\in U} \boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}}$ and (A.31) holds.

Parameters $\boldsymbol{w}$ are now Stage-1 decision variables, so $\mathcal{X}_f$ does not contain adaptive variables. The set of conditions describes a PRO solution to the static robust optimization problem obtained after plugging in decision rule structure $f_{\boldsymbol{w}}(\cdot)$.

A.11 Proof Corollary 5

Corollary 5(i): For any vector of parameters $\boldsymbol{w} \in \mathbb{R}^p$, let $f_{\boldsymbol{w}}(\hat{\boldsymbol{z}})$ denote a decision rule that depends only on $\hat{\boldsymbol{z}} \in \hat{U}$, the non-constraintwise component of uncertain parameter $\boldsymbol{z}$. From Lemma 1 it follows that $\mathcal{X}$ is equal to

$$\mathcal{X}_{\text{hybrid}} = \{\boldsymbol{x} \in \mathbb{R}^{n_x} \mid \exists \boldsymbol{w} \in \mathbb{R}^p : \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B} f_{\boldsymbol{w}}(\hat{\boldsymbol{z}}) \leq \boldsymbol{r}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U\},$$

i.e., the feasible region for $\boldsymbol{x}$ remains unchanged if all adaptive variables are restricted to depend only on the non-constraintwise component of $\boldsymbol{z}$. Hence, setting $X_f = \mathcal{X}_{\text{hybrid}}$ in the proof of Theorem 3 yields the result.

Corollary 5(ii): For each block $v = 1, \ldots, V$, let $\boldsymbol{w}(v) \in \mathbb{R}^{p(v)}$ denote a vector of parameters and let $f^v_{\boldsymbol{w}(v)}(\boldsymbol{z}_{(v)})$ denote a decision rule that depends only on $\boldsymbol{z}_{(v)}$, the uncertain parameters in block v. From Lemma 2 it follows that $\mathcal{X}$ is equal to

$$\mathcal{X}_{\text{block}} = \{\boldsymbol{x} \in \mathbb{R}^{n_x} \mid \forall v = 1, \ldots, V,\ \exists \boldsymbol{w}(v) \in \mathbb{R}^{p(v)} : \boldsymbol{a}_i(\boldsymbol{z}_{(v)})^\top \boldsymbol{x} + \boldsymbol{b}_i^\top f^v_{\boldsymbol{w}(v)}(\boldsymbol{z}_{(v)}) \leq r_i(\boldsymbol{z}_{(v)}), \quad \forall \boldsymbol{z} \in U^v,\ \forall i \in K(v)\},$$

i.e., the feasible region for $\boldsymbol{x}$ remains unchanged if all adaptive variables are restricted to depend only on uncertain parameters in their own block. Hence, setting $X_f = \mathcal{X}_{\text{block}}$ in the proof of Theorem 3 yields the result.

Corollary 5(iii): From Lemma 3 it follows that for simplex uncertainty $\mathcal{X}$ is equal to

$$\mathcal{X}_{\text{simplex}} = \{\boldsymbol{x} \in \mathbb{R}^{n_x} \mid \exists \boldsymbol{u} \in \mathbb{R}^{n_y},\ \boldsymbol{V} \in \mathbb{R}^{n_y \times L} : \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{u} + \boldsymbol{V}\boldsymbol{z}) \leq \boldsymbol{r}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in U\},$$

i.e., the feasible region for $\boldsymbol{x}$ remains unchanged if all adaptive variables are restricted to depend affinely on $\boldsymbol{z}$. Hence, setting $X_f = \mathcal{X}_{\text{simplex}}$ in the proof of Theorem 3 yields the result.

A.12 Proof Lemma 6

The two cases are considered separately.

- *Optimal objective value is zero*: Proof by contradiction. Suppose $\tilde{\boldsymbol{y}}(\cdot)$ is not a PARO extension of $\tilde{\boldsymbol{x}}$. Then, by Definition 8, there exists a $\bar{\boldsymbol{y}}(\cdot)$ such that $(\tilde{\boldsymbol{x}}, \bar{\boldsymbol{y}}(\cdot))$ is ARO to (3) and for some $\bar{\boldsymbol{z}} \in U$ it holds that

$$\boldsymbol{c}(\bar{\boldsymbol{z}})^\top \tilde{\boldsymbol{x}} + \boldsymbol{d}^\top \tilde{\boldsymbol{y}}(\bar{\boldsymbol{z}}) > \boldsymbol{c}(\bar{\boldsymbol{z}})^\top \tilde{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}(\bar{\boldsymbol{z}}).$$

  However, then $(\boldsymbol{z}, \boldsymbol{y}) = (\bar{\boldsymbol{z}}, \bar{\boldsymbol{y}}(\bar{\boldsymbol{z}}))$ is feasible to (13) with positive objective value. This is a contradiction.

- *Optimal objective value is positive*: Let $(\boldsymbol{z^*}, \boldsymbol{y^*})$ denote the optimal solution to (13) and let v^* denote the optimal objective value. The decision rule

$$\boldsymbol{y}(\boldsymbol{z}) = \begin{cases} \tilde{\boldsymbol{y}}(\boldsymbol{z}) & \text{if } \boldsymbol{z} \neq \boldsymbol{z^*} \\ \boldsymbol{y^*} & \text{otherwise,} \end{cases}$$

dominates the decision rule $\tilde{\boldsymbol{y}}(\cdot)$, so the latter is not PARO. We prove the last part by contradiction. Suppose there exists a scenario $\bar{\boldsymbol{z}}$ and a decision $\bar{\boldsymbol{y}}$ such that

$$\begin{aligned} \left(\boldsymbol{c}(\bar{\boldsymbol{z}})^\top \tilde{\boldsymbol{x}} + \boldsymbol{d}^\top \tilde{\boldsymbol{y}}(\bar{\boldsymbol{z}})\right) - \left(\boldsymbol{c}(\bar{\boldsymbol{z}})^\top \tilde{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}\right) &> v^*, \\ \boldsymbol{A}(\bar{\boldsymbol{z}})\tilde{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}} &\leq \boldsymbol{r}(\bar{\boldsymbol{z}}) \end{aligned}$$

i.e., $\bar{\boldsymbol{y}}$ is a feasible wait-and-see decision for scenario $\bar{\boldsymbol{z}}$, and the resulting objective value of $\tilde{\boldsymbol{y}}(\bar{\boldsymbol{z}})$ exceeds that of $\bar{\boldsymbol{y}}$ by more than v^*. Then $(\bar{\boldsymbol{z}}, \bar{\boldsymbol{y}})$ is feasible to (13) with a strictly better objective value than v^*. This is a contradiction.

A.13 Proof Lemma 8

Proof by contradiction, analogous to proof of Theorem 1 of Iancu and Trichakis (2014). Because U is the convex hull of $\boldsymbol{z}^1, \ldots, \boldsymbol{z}^N$, (16b) and (16c) ensure that $\bar{\boldsymbol{x}}$ is ARO to (3) (with $\boldsymbol{d} = \boldsymbol{0}$). Suppose $\bar{\boldsymbol{x}}$ is not PARO to (3). According to Definition 7(i) there exists an $\hat{\boldsymbol{x}}$ that is ARO to (3) and

$$\begin{aligned} \boldsymbol{c}(\boldsymbol{z})^\top \hat{\boldsymbol{x}} &\leq \boldsymbol{c}(\boldsymbol{z})^\top \bar{\boldsymbol{x}}, \quad \forall \boldsymbol{z} \in U, \\ \boldsymbol{c}(\hat{\boldsymbol{z}})^\top \hat{\boldsymbol{x}} &< \boldsymbol{c}(\hat{\boldsymbol{z}})^\top \bar{\boldsymbol{x}}, \quad \text{for some } \hat{\boldsymbol{z}} \in U. \end{aligned}$$

Because $\hat{\boldsymbol{x}}$ is ARO to (3), there also exist $(\hat{\boldsymbol{y}}^{\boldsymbol{1}}, \ldots, \hat{\boldsymbol{y}}^{\boldsymbol{N}})$ that, together with $\hat{\boldsymbol{x}}$, are feasible to (16).

The linear optimization problem $\min_{\boldsymbol{z} \in U} \boldsymbol{c}(\boldsymbol{z})^\top (\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}})$ attains the minimum in a vertex solution, so without loss of generality we can assume $\hat{\boldsymbol{z}} \in \text{ext}(U)$. Any point $\bar{\boldsymbol{z}} \in \text{ri}(U)$ can be written as a strict convex combination of the extreme points of U (Rockafellar, 1970), so $\bar{\boldsymbol{z}} = \sum_{i=1}^N \alpha_i \boldsymbol{z}^i$ for some $\boldsymbol{\alpha} \in \mathbb{R}^N$ with $\sum_{i=1}^N \alpha_i = 1$, $\alpha_i > 0$ for all i. Then

$$\boldsymbol{c}(\bar{\boldsymbol{z}})^\top (\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}}) = \sum_{\substack{i=1 \\ \boldsymbol{z}^i \neq \hat{\boldsymbol{z}}}}^N \alpha_i \boldsymbol{c}(\boldsymbol{z}^i)^\top (\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}}) + \hat{\alpha} \boldsymbol{c}(\hat{\boldsymbol{z}})^\top (\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}}),$$

where the first term of the RHS is nonpositive and the second term is strictly negative. This contradicts the fact that $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}^{\boldsymbol{1}}, \ldots, \bar{\boldsymbol{y}}^{\boldsymbol{N}})$ is optimal to (16).

A.14 Proof Lemma 9

First, we note that in problem $p(\hat{\boldsymbol{x}}, V)$, setting Stage-1 decision $\bar{\boldsymbol{x}} = \hat{\boldsymbol{x}}$ is feasible, and in that case for any scenario the optimal recourse decision is the same for the original $\bar{\boldsymbol{x}}$ and the new $\bar{\boldsymbol{x}}$. Thus, $p(\hat{\boldsymbol{x}}, V)$ is always nonpositive.

Suppose Algorithm 1 terminates after k iterations with solution $\bar{\boldsymbol{x}}$. Then $\bar{\boldsymbol{x}}$ is part of the optimal solution to problem $p(\boldsymbol{x}^{\boldsymbol{k-1}}, V_k)$ which is solved in iteration k and yields objective value 0. Thus, due to (17c) and (17d) and the fact that the extreme points are in set V_k, we know that $\bar{\boldsymbol{x}}$ is ARO.

It remains to show that $\bar{\boldsymbol{x}}$ is PARO according to Definition 7(i). Proof by contradiction. Suppose $\bar{\boldsymbol{x}}$ is not PARO. Then there exists another $\boldsymbol{x^*}$ that is ARO to (3) that additionally satisfies the following two conditions:

1. For each $\boldsymbol{z} \in U$ there exists a $\boldsymbol{y}$ such that for all $\bar{\boldsymbol{y}}$ with $\boldsymbol{A}\bar{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}} \leq \boldsymbol{r}(\boldsymbol{z})$ we have

$$\begin{aligned} \boldsymbol{c}^\top \boldsymbol{x^*} + \boldsymbol{d}^\top \boldsymbol{y} &\leq \boldsymbol{c}^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}, \\ \boldsymbol{A}\boldsymbol{x^*} + \boldsymbol{B}\boldsymbol{y} &\leq \boldsymbol{r}(\boldsymbol{z}). \end{aligned}$$

2. There exists a $\boldsymbol{z^*} \in U$ and a $\boldsymbol{y^*}$ such that for all $\bar{\boldsymbol{y}}$ with $\boldsymbol{A}\bar{\boldsymbol{x}} + \boldsymbol{B}\bar{\boldsymbol{y}} \leq \boldsymbol{r}(\boldsymbol{z^*})$ we have

$$\begin{aligned} \boldsymbol{c}^\top \boldsymbol{x^*} + \boldsymbol{d}^\top \boldsymbol{y^*} &< \boldsymbol{c}^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}, \\ \boldsymbol{A}\boldsymbol{x^*} + \boldsymbol{B}\boldsymbol{y^*} &\leq \boldsymbol{r}(\boldsymbol{z^*}). \end{aligned}$$

Because for every $(\boldsymbol{z}^i, v^i) \in V_k$ it holds that $\boldsymbol{z}^i \in U$, the first condition implies that for each $(\boldsymbol{z^i}, v^i) \in V^k$ there exists a recourse decision $\boldsymbol{y^{i*}}$ such that $(\boldsymbol{x^*}, \boldsymbol{y^{i*}})$ satisfies constraints (17c) and (17d). The second condition is equivalent to the statement that there exists $\boldsymbol{z^*} \in U$ and a $\boldsymbol{y^*}$ such that

$$\max_{\bar{\boldsymbol{y}}:\boldsymbol{A}\bar{\boldsymbol{x}}+\boldsymbol{B}\bar{\boldsymbol{y}}\leq \boldsymbol{r}(\boldsymbol{z^*})} (\boldsymbol{c}^\top \boldsymbol{x^*} + \boldsymbol{d}^\top \boldsymbol{y^*}) - (\boldsymbol{c}^\top \bar{\boldsymbol{x}} + \boldsymbol{d}^\top \bar{\boldsymbol{y}}) < 0,$$
$$\boldsymbol{A}\boldsymbol{x^*} + \boldsymbol{B}\boldsymbol{y^*} \leq \boldsymbol{r}(\boldsymbol{z^*}).$$

Put together, this implies that $(\boldsymbol{z^*}, \boldsymbol{x^*}, \boldsymbol{y^*}, \boldsymbol{y^{1*}}, \ldots, \boldsymbol{y^{|V_k|*}})$ is a feasible solution to (17) with strictly negative objective value. This contradicts with $p(\boldsymbol{x^{k-1}}, V_k) = 0$. Thus, $\bar{\boldsymbol{x}}$ is PARO.