

A Unified Framework for Adjustable Robust Optimization with Endogenous Uncertainty

Qi Zhang ^{*1} and Wei Feng²

¹*Department of Chemical Engineering and Materials Science,
University of Minnesota, Minneapolis, MN 55455, USA*

²*State Key Laboratory of Industrial Control Technology, College of Control Science and
Engineering, Zhejiang University, Hangzhou 310027, China*

Abstract

This work proposes a framework for multistage adjustable robust optimization that unifies the treatment of three different types of endogenous uncertainty, where decisions, respectively, (i) alter the uncertainty set, (ii) affect the materialization of uncertain parameters, and (iii) determine the time when the true values of uncertain parameters are observed. We provide a systematic analysis of the different types of endogenous uncertainty and highlight the connection between optimization under endogenous uncertainty and active learning. We consider decision-dependent polyhedral uncertainty sets and propose a decision rule approach that incorporates both continuous and binary recourse, including recourse decisions that affect the uncertainty set. The proposed method enables the modeling of decision-dependent nonanticipativity and results in a tractable reformulation of the problem. We demonstrate the effectiveness of the approach in computational experiments that cover a range of applications, including plant redesign, maintenance planning with inspections, optimizing revision points in capacity planning, and production scheduling with active parameter estimation. The results show significant benefits from the proper modeling of endogenous uncertainty and active learning.

Keywords: adjustable robust optimization, endogenous uncertainty, active learning

1 Introduction

Robust optimization is an approach to decision making under uncertainty that has gained tremendous popularity in the optimization community over the last two decades. It is particularly attractive in situations in which (i) the set of possible uncertainty realizations, commonly referred to as the uncertainty set, is known but the probability distribution is not, (ii) feasibility over the entire

^{*}Corresponding author (qizh@umn.edu)

uncertainty set or the worst-case performance is of main interest, or (iii) alternative approaches, such as scenario-based stochastic programming, are computationally significantly less tractable. Earlier works on robust optimization^{1,2} were limited to the *static* case, which only considers *here-and-now* decisions that need to be determined before the uncertainty is realized. In contrast, *adjustable* robust optimization (ARO)³ provides a framework that also incorporates *wait-and-see* (or *recourse*) decisions, which can be determined after the true values of some uncertain parameters are revealed. This makes ARO suitable for a large set of sequential decision-making problems, hence greatly expanding the applicability of the robust optimization paradigm. There has been a plethora of theoretical and practical contributions addressing a variety of ARO problems. A particular research focus has been the development of tractable approximations, e.g. in the form of decision rules^{3,4}, and solution methods^{5,6}, with many of the more recent efforts directed at the efficient handling of discrete recourse decisions^{7–12}. We refer the reader to Yanıkoğlu et al.¹³ for a comprehensive survey on ARO.

The term “adjustable robust optimization” had not appeared in the process systems engineering (PSE) literature until 2016^{14–16}. Only then, as formally discussed by Zhang et al.¹⁷, the PSE community realized that it actually has been applying the concept of (two-stage) ARO for decades, but under a different name: flexibility analysis¹⁸. Many of the approaches developed in this line of work resemble the ones presented in the robust optimization literature. However, traditional flexibility analysis focuses primarily on nonlinear problems, which may be the reason why it so far has not been recognized or noticed by the operations research community. In the last few years, the number of ARO-related works in PSE has increased rapidly, addressing diverse applications in process design^{19,20}, planning and scheduling^{14–16,21–24}, model predictive control^{25–27}, supply chain optimization^{28,29}, etc.

In stochastic optimization, one distinguishes between *exogenous* and *endogenous* uncertainty. While exogenous uncertainty is not affected by the decision maker’s actions, certain properties of endogenous uncertain parameters are decision-dependent. Endogenous uncertainty has its origin in stochastic programming³⁰, and mainly two types of endogenous uncertainty have been considered in the literature: *type 1* where decisions alter the underlying probability distribution of an uncertain parameter^{31–33}, and *type 2* where decisions affect the time at which an uncertain parameter materializes or its true value is revealed. The vast majority of existing works address type-2 endogenous uncertainty^{34–43}. Here, the main challenge is the modeling of decision-dependent nonanticipativity, which is typically achieved by encoding a conditional scenario tree. The size of this scenario tree grows exponentially with the number of uncertain parameters and decisions affecting the uncertainty, which leads to dramatically increased computational complexity compared to the case with only exogenous uncertainty. Hence, most research efforts have focused on the development of more efficient formulations^{35,36,38,42,43} and solution methods^{35,37,39–41,43}. We refer to Apap and Grossmann⁴³ for a comprehensive review of stochastic programming with endogenous uncertainty.

Only recently, endogenous uncertainty has also been considered in robust optimization. Poss⁴⁴ and Nohadani and Sharma⁴⁵ address static robust optimization with decision-dependent polyhedral uncertainty sets. A similar approach is taken by Lappas and Gounaris⁴⁶, who further extend their framework to consider multistage robust process scheduling where the materialization of task-specific uncertainties, such as processing time and production yield, depends on the

execution of the task^{14,22}; however, the uncertainty set is only affected by first-stage decisions, and all recourse variables are continuous. Avraamidou and Pistikopoulos⁴⁷ apply multiparametric programming to address endogenous uncertainty in two-stage robust optimization. Finally, Feng et al.⁴⁸ consider the multistage case with both continuous and binary recourse as well as uncertainty sets that can be affected by decisions at every stage, for which a decision rule approach is applied to derive tractable approximations.

The above-mentioned references investigate problems in which decisions directly affect the stochastic nature of uncertain parameters, which in the robust optimization context translates into uncertainty sets whose shape, size, or dimensionality may be decision-dependent. Also, although the materialization of uncertain parameters can be decision-dependent, it is assumed that the true values of uncertain parameters are observed immediately after their materialization. In a series of recent works, Vayanos and coworkers^{49–51} consider robust optimization with decision-dependent information discovery³⁸, where decisions determine whether and when uncertain parameters are observed. This constitutes a conceptually different type of endogenous uncertainty, where, instead of an underlying stochastic process, the source of uncertainty is the lack of information, which can be acquired with additional effort. In this case, the uncertainty set is reduced (“shrinks”) with additional observations. The proposed framework provides a means of modeling *active learning*, which refers to selecting experiments sequentially based on the outcomes of previous experiments. It is hence a promising approach to optimizing the trade-off between *exploration* and *exploitation* in sequential decision-making problems, particularly those involving complex constraints. Existing works have applied this approach to the learning of customer behavior through dynamic pricing⁴⁹ and active preference elicitation with application to housing allocation⁵¹.

We have come to realize that many problems in PSE require the simultaneous consideration of multiple types of endogenous uncertainty. Examples include strategic capacity planning, design of sensor networks, integrated online process optimization and parameter estimation, and maintenance planning with inspections. To address these problems, we need an approach that can consider all previously described types of endogenous uncertainty simultaneously, which, to the best of our knowledge, does not yet exist in the literature. In this work, we develop such a unified ARO framework, which significantly expands our capability to model data-driven optimization problems under uncertainty, particularly those involving active learning. The main contributions of this work are as follows:

- We introduce a new refined classification of endogenous uncertainty that more comprehensively reflects the current understanding of the subject. We further provide a unifying robust optimization perspective and highlight the connection between endogenous uncertainty and active learning.
- We present two-stage and multistage robust optimization formulations and show that they can be used to simultaneously consider all defined types of endogenous uncertainty subject to polyhedral uncertainty sets.
- A decision rule approach based on previous work⁴⁸ is employed to develop tractable approximations of the given problems, which involve both continuous and binary recourse variables. One major novelty of this work is the use of auxiliary uncertain parameters to

model decision-dependent nonanticipativity in the multistage setting. We derive a reformulation that can be directly solved using off-the-shelf optimization solvers.

- The versatility and efficacy of the proposed modeling framework are demonstrated in several computational case studies, which include plant redesign, maintenance planning with inspections, optimizing revision points in capacity planning, and production scheduling with active parameter estimation.

The remainder of this paper is organized as follows. In Section 2, we propose a new refined classification of endogenous uncertainty and systematically analyze the different types of endogenous uncertainty from a robust optimization perspective. In Section 3, the connection between endogenous uncertainty and active learning is established. We develop the two-stage formulation incorporating all types of endogenous uncertainty in Section 4 before presenting the multistage formulation in Section 5. The proposed decision rule approach is detailed in Section 6. In Section 7, results from the computational studies are presented. Finally, in Section 8, we close with some concluding remarks.

Notation We use lowercase and uppercase boldface letters to denote vectors and matrices, respectively, e.g. $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n \times m}$. Scalar quantities are denoted by non-boldface letters. A vector \mathbf{x} 's i th element is denoted by x_i . We use \circ and $\mathbb{1}(\cdot)$ to denote the Hadamard multiplication operator and the indicator function, respectively. Furthermore, $\mathbf{0}$, \mathbf{e} , and \mathbf{E} are the zero vector, all-ones vector, and all-ones matrix, respectively, while \mathbf{e}_i is the standard basis vector whose i th element is 1; their dimensions can be inferred from the context.

2 Endogenous Uncertainty Through the Robust Optimization Lens

As mentioned in Section 1, the literature distinguishes between type-1 and type-2 endogenous uncertainty. Here, we further refine the type-2 classification and propose the following definition of *three* different types of endogenous uncertainty:

- Type 1: Decisions alter the probability distribution of uncertain parameters.
- Type 2a: Decisions determine whether and when uncertain parameters materialize, i.e. actually become physically meaningful.
- Type 2b: Decisions determine whether and when the true values of uncertain parameters are observed.

Endogenous uncertain parameters can be of more than one type. For example, a company may alter the probability distribution of the demand for its product by changing the selling price, in which case the product demand is a type-1 endogenous uncertain parameter. In addition, if we still have to decide whether or not to launch the product in the first place, the product demand is also type-2a endogenous since it will only materialize if the product is actually available on the market. Moreover, a customer's product demand is observed when the order is placed; it is type-2b endogenous if we can affect the time of the order, e.g. by specifying it in a contract.

In robust optimization, instead of specifying a probability distribution, uncertainty is described using an uncertainty set, which is the set of all realizations of the uncertain parameters considered possible in the problem. In other words, robust optimization only requires the support of the probability distribution. A two-stage robust optimization problem can be generally formulated as follows:

$$\begin{aligned} & \underset{x \in \mathcal{X}, \tilde{x}}{\text{minimize}} && \max_{\xi \in \Xi} f(x, \tilde{x}(\xi), \xi) \\ & \text{subject to} && g(x, \tilde{x}(\xi), \xi) \leq \mathbf{0}, \tilde{x}(\xi) \in \tilde{\mathcal{X}} \quad \forall \xi \in \Xi, \end{aligned} \quad (1)$$

where $x \in \mathcal{X}$ are the first-stage variables and $\tilde{x} \in \tilde{\mathcal{X}}$ are the second-stage variables, which are functions of the uncertain parameters $\xi \in \mathbb{R}^K$. The objective and constraint functions are denoted by f and g , respectively. Problem (1) minimizes the worst-case cost while ensuring that the constraints are satisfied for all $\xi \in \Xi$. By applying an epigraph reformulation, a problem of form (1) can be reformulated into a problem of the following form:

$$\begin{aligned} & \underset{x \in \mathcal{X}, \tilde{x}}{\text{minimize}} && x_1 \\ & \text{subject to} && g(x, \tilde{x}(\xi), \xi) \leq \mathbf{0}, \tilde{x}(\xi) \in \tilde{\mathcal{X}} \quad \forall \xi \in \Xi, \end{aligned} \quad (2)$$

where the objective function is deterministic. Formulation (2) will be used in the remainder of this paper as it is more convenient for our discussion and subsequent reformulations.

Remark There is a common misconception that a robust optimization problem has to consider a worst-case cost function. As it is apparent from (2), in theory, one can also consider any deterministic or deterministic equivalent (e.g. expressed in terms of scenarios) cost function. Constraint satisfaction with respect to the entire uncertainty set can be considered independently from the choice of objective function. Often, a non-worst-case objective function is more appropriate (see Section 7 for two examples).

In the case of endogenous uncertainty in the two-stage setting, the uncertainty depends on the first-stage decisions x . For each type of endogenous uncertainty, we show how x affect the following three major components of problem (2):

- the uncertainty set Ξ , which is constant in the case of only exogenous uncertainty but depends on x if the uncertainty is endogenous, i.e. $\Xi = \Xi(x)$;
- the *deterministic feasible set* (DFS) defined as

$$\mathcal{F}(x, \xi) := \left\{ \tilde{x} \in \tilde{\mathcal{X}} : g(x, \tilde{x}, \xi) \leq \mathbf{0} \right\},$$

which is the set of feasible second-stage solutions given x and ξ ;

- and the second-stage recourse decisions $\tilde{x}(\xi)$ as functions of ξ .

Type 1 (uncertainty alteration) The probability distribution of type-1 endogenous uncertain parameters including its support can be altered by decisions. From the robust optimization perspective, this means that the shape and size of the uncertainty set $\Xi(x)$ depend on the first-stage

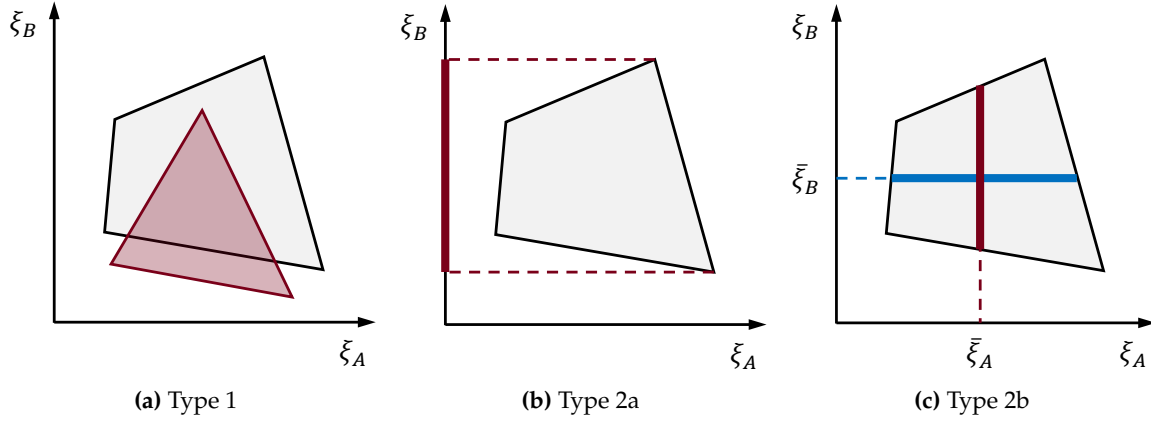


Figure 1: Decisions affect the uncertainty set differently depending on the type of endogenous uncertainty: (a) type 1: shape and size of the uncertainty set may be altered; (b) type 2a: dimensionality of the uncertainty set changes depending on which uncertain parameters materialize; (c) type 2b: uncertainty set does not change, but its reduction over time does, i.e. it only reduces if realization of uncertainty is observed.

decisions x . For an example involving two uncertain parameters, ξ_A and ξ_B , Figure 1a shows the uncertainty sets for two different choices of x . One results in a polyhedral uncertainty set with four facets while the other results in a polytope with three facets.

Type 2a (uncertainty materialization) In the case of type-2a endogenous uncertainty, x change the uncertainty set by affecting which uncertain parameters actually materialize. In the example shown in Figure 1b, $\Xi(x)$ is the gray shaded polytope if both ξ_A and ξ_B materialize. However, if x are chosen such that ξ_A does not materialize, $\Xi(x)$ becomes the solid red line segment that is a one-dimensional set for ξ_B ; hence, x affect the dimensionality of the uncertainty set. Generally, when some uncertain parameters do not materialize, $\Xi(x)$ becomes the projection of the full-dimensional uncertainty set onto the space of the materialized uncertain parameters.

If an uncertain parameter does not materialize, it is physically meaningless and should therefore be irrelevant for the problem. More precisely, this means that the following DFS dependence condition has to hold: if x are chosen such that some uncertain parameters do not materialize, then $\mathcal{F}(x, \xi)$ has to be independent of those unmaterialized parameters. It is important to note that decision-dependent materialization of parameters is not a feature of uncertainty. It equally occurs in deterministic problems. Hence, the deterministic model, in which ξ is fixed, should already satisfy the DFS dependence condition, which usually directly extends to the robust formulation.

Finally, recourse decisions cannot depend on unmaterialized parameters. This, however, does not imply that recourse decisions are functions of all materialized parameters. While materialization is a necessary condition for an uncertain parameter to be considered in the recourse function, the sufficient condition is that its true value is observed. This subtle but important distinction is often ignored as almost all existing works on endogenous uncertainty assume that once an uncertain parameter materializes, its true value is automatically observed.

Type 2b (uncertainty observation) In the case of type-2b endogenous uncertainty, we can decide if and when (in the multistage case) the true values of uncertain parameters are observed. Here, the underlying probability distribution and hence the uncertainty set is unaffected, but the reduction of uncertainty over time depends on when and which observations are made. This is illustrated in Figure 1c, which shows an instance in which the true values of the uncertain parameters ξ_A and ξ_B are $\bar{\xi}_A$ and $\bar{\xi}_B$, respectively. If both parameters are observed, there will be no uncertainty in the second stage as the uncertainty set reduces to the point $(\bar{\xi}_A, \bar{\xi}_B)$. However, if we only observe ξ_A and find it to be $\bar{\xi}_A$, the uncertainty with respect to ξ_B remains and its reduced uncertainty set becomes the red solid line. Similarly, if we only observe ξ_B , the reduced uncertainty set for ξ_A becomes the blue solid line. Note that the reduced uncertainty set is conditional, i.e. it depends on the observations made. Clearly, a parameter can only be observed if it materializes. Also, reiterating our previous remark, recourse decisions can only be functions of observed uncertain parameters.

The above discussion references the two-stage robust optimization problem but naturally extends to the multistage case. An overview of the characteristics of the different types of endogenous uncertainty is shown in Table 1.

Table 1: Overview of the impact of decisions on the uncertainty set, the dependence of the DFS on uncertain parameters, and the dependence of the recourse functions on uncertain parameters for the different types of endogenous uncertainty.

	Uncertainty Set	DFS Dependence	Recourse Dependence
Type 1	shape, size	unaffected	unaffected
Type 2a	dimensionality	only dependence on materialized parameters	no dependence on unmaterialized parameters
Type 2b	reduction over time	unaffected	only dependence on observed parameters

3 Endogenous Uncertainty and Active Learning

In the context of statistical learning, *active learning* is the concept of actively gathering information for the purpose of statistical inference. It is also referred to as optimal experimental design as the decisions regarding information discovery are made with an objective in mind, such as minimizing the number of data points required to achieve a given level of confidence in the resulting statistical model. In this section, we highlight the connection between active learning and endogenous uncertainty, and show how it can be conceptualized in a set-based robust optimization framework.

The goal of learning is to gain insights into unknown properties, which can be viewed as uncertain parameters whose level of uncertainty we wish to reduce. Hence, learning is a means of uncertainty reduction. It does so by making new observations that *condition* the probability distribution; it does not *alter* the underlying distribution, which is an important distinction. With this perspective, we can model active learning using the notion of endogenous uncertainty. Let

$\xi \in \mathbb{R}^K$ be the uncertain parameters of interest and $\zeta \in \mathbb{R}^M$ be uncertain parameters that can be observed before the realization of ξ . Before observing any ζ , the probability distribution of ξ is $\mathbb{P}(\xi)$. The decision maker can now choose which ζ to observe in order to improve the current information state with regard to ξ , indicated by the binary variables $z \in \{0, 1\}^M$ where z_i equals 1 if ζ_i is observed and 0 otherwise. After observing the chosen parameters, the updated probability distribution becomes $\mathbb{P}(\xi \mid \zeta_i = \hat{\zeta}_i \ \forall i \text{ for which } z_i = 1)$, where $\hat{\zeta}$ denotes a specific observed realization of ζ . In a Bayesian sense, $\mathbb{P}(\xi)$ and $\mathbb{P}(\xi|\zeta)$ can be interpreted as the prior and posterior distributions, respectively. Clearly, in this setting, ζ are type-2b endogenous uncertain parameters. They may also be type-2a endogenous if their materialization is also decision-dependent. Note that ξ are considered exogenous uncertain parameters if their materialization and observation are not decision-dependent and if $\mathbb{P}(\xi)$ and $\mathbb{P}(\xi|\zeta)$ are fixed.

In a robust optimization setting, uncertainty reduction implies reducing the uncertainty set. Using the same notation as above, uncertainty reduction through active learning can then be characterized as follows: Let Ω be the uncertainty set for the uncertain parameters (ξ, ζ) before the observation of any ζ , i.e. $(\xi, \zeta) \in \Omega$. Once some ζ are observed (indicated by the binary variables z) with the realization $\hat{\zeta}$, the uncertainty set reduces to $\bar{\Omega}(z, \hat{\zeta}) = \{(\xi, \zeta) : (\xi, \zeta) \in \Xi, \ \zeta = z \circ \hat{\zeta}\}$. Note that $\bar{\Omega}(z, \hat{\zeta}) \subseteq \Omega$ for every possible $\hat{\zeta}$ as specified by Ω .

Example 1. We consider investing in a large-scale industrial manufacturing plant, where the real plant capacity q is unknown before it is actually built. To reduce the financial risk, we have the option of building a small-scale pilot plant whose achieved capacity p can help better predict the capacity of the industrial-scale plant. The uncertainty set depicted in Figure 2 shows that p and q are highly correlated. If no pilot plant is built, the uncertain parameter p does not materialize. As a result, we can only rely on our initial belief that the marginal uncertainty set for q is $\mathcal{Q} = [q^{\min}, q^{\max}]$. However, if we do build a pilot plant and observe a pilot plant capacity of $\bar{p} \in [p^{\min}, p^{\max}]$, the uncertainty set for q reduces to a significantly smaller $\bar{\mathcal{Q}}(\bar{p}) = \{q : \alpha + \beta\bar{p} - \Delta \leq q \leq \alpha + \beta\bar{p} + \Delta\}$. Note that in this example, p is a type-2a and type-2b uncertain parameter since it only materializes (and can then be observed) if we decide to build the pilot plant.

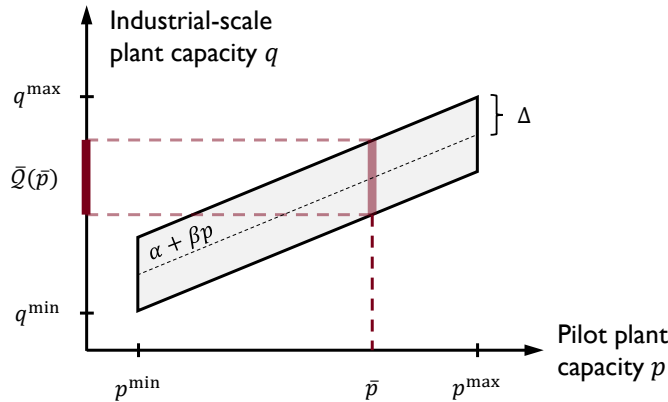


Figure 2: The uncertainty set for the industrial-scale plant capacity is reduced by observing the capacity of a pilot plant.

In Example 1, we use the term *marginal uncertainty set* and define it as the projection of the

full-dimensional uncertainty set onto the space of a subset of variables, in this case q . Generally, given uncertain parameters ξ and ζ , observing ζ further restricts the marginal uncertainty set for ξ , denoted by Ξ , only if the uncertain parameters are correlated. The geometric interpretation is depicted in Figure 3, where an observation $\hat{\zeta}$ results in cutting planes (indicated by the red dashed lines) added to Ξ such that the updated marginal uncertainty set for ξ becomes $\bar{\Xi}(\hat{\zeta})$.

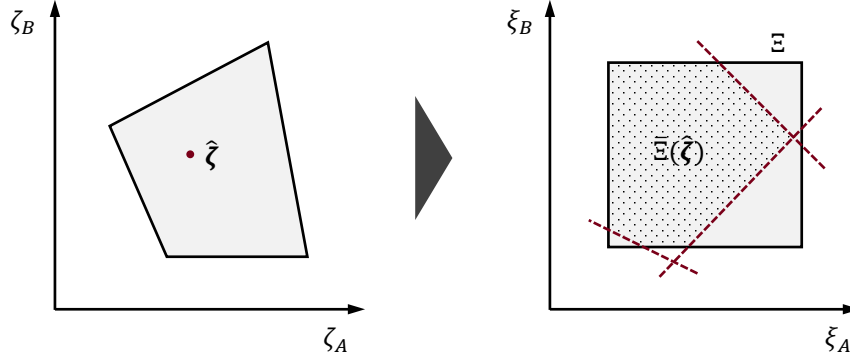


Figure 3: The marginal uncertainty set for ξ is reduced from Ξ (shaded rectangle) to $\bar{\Xi}(\hat{\zeta})$ (pattern-filled polytope) by added cutting planes corresponding to observation $\hat{\zeta}$.

Modeling active learning as a robust optimization problem with endogenous uncertainty provides an opportunity for integrated optimization and learning, which are two tasks typically performed separately. Central to this problem is the trade-off between exploration and exploitation since our actions affect both the amount of new information and the reward that we receive. The proposed framework can readily incorporate this trade-off and is particularly well suited for problems with complex constraints and recourse decisions, which often cause difficulties for many alternative methods, such as Bayesian optimization⁵², reinforcement learning⁵³, and multi-armed bandit optimization⁵⁴.

4 Two-Stage Robust Optimization with Endogenous Uncertainty

In the following, we show how type-1, type-2a, and type-2b endogenous uncertainty can be incorporated in a two-stage robust optimization formulation, eventually leading to a unified model that can capture all these types of endogenous uncertainty simultaneously.

4.1 Incorporating Type-1 Endogenous Uncertainty

Consider the following two-stage robust optimization problem:

$$\begin{aligned}
 & \underset{x, \tilde{x}}{\text{minimize}} && x_1 \\
 & \text{subject to} && x \in \mathcal{X} \\
 & && \left. \begin{aligned} A(\xi)x + \tilde{A}(\xi)\tilde{x}(\xi) &\leq b(\xi) \\ \tilde{x}(\xi) &\in \tilde{\mathcal{X}} \end{aligned} \right\} \quad \forall \xi \in \Xi(x),
 \end{aligned} \tag{3}$$

where the dependence of the uncertainty set Ξ on x indicates type-1 endogenous uncertainty. Throughout this work, we restrict our discussion to decision-dependent uncertainty sets of the following polyhedral form:

$$\Xi(x) = \{\xi \in \mathbb{R}^K : W\xi \leq Ux\}, \quad (4)$$

where we assume that $\Xi(x)$ is always nonempty and bounded, i.e. $\Xi(x) \subseteq \{\xi \in \mathbb{R}^K : \xi^{\min} \leq \xi \leq \xi^{\max}\}$ for any $x \in \mathcal{X}$.

Remark Some endogenous uncertainty can be modeled as exogenous uncertainty through a simple transformation. For example, consider a scalar endogenous uncertain parameter ξ with the uncertainty set $\Xi(x) = \{\xi \in \mathbb{R} : 0 \leq \xi \leq x\}$, which depends on variable x . In this case, we could define a new uncertain parameter $\tilde{\xi}$ and add the equation $\xi = \tilde{\xi}x$ to the model. The uncertainty set for $\tilde{\xi}$ is then $\tilde{\Xi} = \{\tilde{\xi} \in \mathbb{R} : 0 \leq \tilde{\xi} \leq 1\}$, which does not depend on x , and ξ can now be treated as a variable that is adjustable with respect to $\tilde{\xi}$.

4.2 Incorporating Type-2a Endogenous Uncertainty

To model type-2a endogenous uncertainty, we introduce binary variables $y \in \{0, 1\}^K$ to encode the materialization of uncertain parameters, i.e. $y_i = 1$ if and only if ξ_i materializes. Then, both type-1 endogenous uncertainty with decision-dependent uncertainty set of the form (4) and type-2a endogenous uncertainty can be captured in a two-stage formulation of the following form:

$$\begin{aligned} & \underset{x, y, \tilde{x}}{\text{minimize}} && x_1 \\ & \text{subject to} && x \in \mathcal{X}, y \in \mathcal{Y} \\ & && \left. \begin{aligned} A(\xi)x + D(\xi)y + \tilde{A}(\xi)\tilde{x}(y \circ \xi) &\leq b(\xi) \\ \tilde{x}(y \circ \xi) &\in \tilde{\mathcal{X}} \end{aligned} \right\} \quad \forall \xi \in \Xi(x, y), \end{aligned} \quad (5)$$

where $\mathcal{Y} \subseteq \{0, 1\}^K$ and

$$\Xi(x, y) = \{\xi \in \mathbb{R}^K : W\xi \leq Ux + \bar{U}y\}, \quad (6)$$

assuming that $\Xi(x, y) \subseteq \{\xi \in \mathbb{R}^K : \xi^{\min} \leq \xi \leq \xi^{\max}\}$ for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Note that formulation (5) is actually of the same form as (3); we have merely augmented the first-stage variables with y and restricted the recourse functions to only depend on materialized parameters, i.e. $y \circ \xi$. Here, the assumption is that all materialized uncertain parameters are also observed.

While the recourse dependence property of type-2a endogenous uncertainty is explicitly stated, it may not be immediately obvious how the change of dimensionality of the uncertainty set and the DFS dependence property can be captured in (5). Here, the uncertainty set $\Xi(x, y)$ has to be constructed such that for a given y , its projection onto the space of materialized uncertain parameters, i.e. ξ_i for which $y_i = 1$, forms the desired uncertainty set for the given y . In addition, the constraints in (5) have to be constructed such that the feasible region is not affected by any $\xi_i \in \Xi_i(x, y)$ for which $y_i = 0$, where $\Xi_i(x, y)$ denotes the projection of $\Xi(x, y)$ onto the ξ_i -space. Uncertainty sets and constraints satisfying these requirements can be constructed in various ways. To show the generality of formulation (5), we provide the following result.

Proposition 1. *Formulation (5) is general in a sense that it can model the general case in which every possible set of materialized uncertain parameters results in a distinct set of constraints and a distinct x -dependent uncertainty set of appropriate dimensionality.*

Proof. We provide a constructive proof for Proposition 1. In the following, we refer to a possible set of materialized uncertain parameters as a *materialization scenario*. Since $\mathbf{y} \in \{0, 1\}^K$ is defined such that each materialization scenario is given by a specific \mathbf{y} with $y_i = 1$ if and only if the uncertain parameter ξ_i materializes, the set of materialization scenarios is finite (maximum 2^K). Let S be the number of materialization scenarios and $\mathcal{S} := \{1, \dots, S\}$. We define a set $\mathcal{Y} := \{\bar{\mathbf{y}}^1, \dots, \bar{\mathbf{y}}^S\}$ where $\bar{\mathbf{y}}^s$ is the specific \mathbf{y} -vector that represents scenario s . In the most general case, each materialization scenario is associated with a distinct set of constraints and a distinct uncertainty set involving only the materialized uncertain parameters. Hence, the two-stage problem can be formulated as follows:

$$\begin{aligned} & \underset{\mathbf{x}, \mathbf{y}, \tilde{\mathbf{x}}}{\text{minimize}} && x_1 \\ & \text{subject to} && \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y} \\ & && \bigvee_{s \in \mathcal{S}} \left[\begin{array}{c} \mathbf{y} = \bar{\mathbf{y}}^s \\ \mathbf{A}^s(\boldsymbol{\xi}^s) \mathbf{x} + \mathbf{D}^s(\boldsymbol{\xi}^s) \bar{\mathbf{y}}^s + \tilde{\mathbf{A}}^s(\boldsymbol{\xi}^s) \tilde{\mathbf{x}}(\boldsymbol{\xi}^s) \leq \mathbf{b}^s(\boldsymbol{\xi}^s) \\ \tilde{\mathbf{x}}(\boldsymbol{\xi}^s) \in \tilde{\mathcal{X}} \end{array} \right] \forall \boldsymbol{\xi}^s \in \Xi^s(\mathbf{x}) \end{aligned} \quad (7)$$

where the disjunction indicates that only the constraints for the selected materialization scenario need to be satisfied. Here, $\boldsymbol{\xi}^s$ only contains the ξ_i for which $\bar{y}_i^s = 1$ and is therefore a K^s -dimensional vector with $K^s = \mathbf{e}^\top \bar{\mathbf{y}}^s$. The K^s -dimensional scenario-specific uncertainty set is

$$\Xi^s(\mathbf{x}) = \{\boldsymbol{\xi}^s \in \mathbb{R}^{K^s} : \mathbf{W}^s \boldsymbol{\xi}^s \leq \mathbf{U}^s \mathbf{x}\} \subseteq \{\boldsymbol{\xi}^s \in \mathbb{R}^{K^s} : \xi_i^{\min} \leq \xi_i \leq \xi_i^{\max} \ \forall i \text{ for which } \bar{y}_i^s = 1\}.$$

Similar to the constraints, a disjunction can also be used to define an overall uncertainty set $\Xi(\mathbf{x}, \mathbf{y})$. By doing so, problem (7) can be reformulated as

$$\begin{aligned} & \underset{\mathbf{x}, \mathbf{y}, \tilde{\mathbf{x}}}{\text{minimize}} && x_1 \\ & \text{subject to} && \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y} \\ & && \bigvee_{s \in \mathcal{S}} \left[\begin{array}{c} \mathbf{y} = \bar{\mathbf{y}}^s \\ \mathbf{A}^s(\boldsymbol{\xi}^s) \mathbf{x} + \mathbf{D}^s(\boldsymbol{\xi}^s) \bar{\mathbf{y}}^s + \tilde{\mathbf{A}}^s(\boldsymbol{\xi}^s) \tilde{\mathbf{x}}(\bar{\mathbf{y}}^s \circ \boldsymbol{\xi}) \leq \mathbf{b}^s(\boldsymbol{\xi}^s) \\ \tilde{\mathbf{x}}(\bar{\mathbf{y}}^s \circ \boldsymbol{\xi}) \in \tilde{\mathcal{X}} \end{array} \right] \forall \boldsymbol{\xi} \in \Xi(\mathbf{x}, \mathbf{y}) \end{aligned} \quad (8)$$

with

$$\Xi(\mathbf{x}, \mathbf{y}) = \left\{ \boldsymbol{\xi} \in \mathbb{R}^K : \bigvee_{s \in \mathcal{S}} \left[\begin{array}{c} \mathbf{y} = \bar{\mathbf{y}}^s \\ \mathbf{W}^s \boldsymbol{\xi}^s \leq \mathbf{U}^s \mathbf{x} \\ \xi_i = 0 \ \forall i \text{ for which } \bar{y}_i^s = 0 \end{array} \right] \right\}.$$

Note that in $\Xi(\mathbf{x}, \mathbf{y})$, we set unmaterialized uncertain parameters to zero; this is a convenient but arbitrary choice since for a given materialization scenario s , only the projection of $\Xi(\mathbf{x}, \mathbf{y})$ onto the $\boldsymbol{\xi}^s$ -space matters.

With additional binary variables $\omega \in \{0, 1\}^S$, we can reformulate the disjunctions and arrive at the following formulation:

$$\begin{aligned}
& \underset{\mathbf{x}, \omega, \mathbf{y}, \tilde{\mathbf{x}}}{\text{minimize}} && x_1 \\
& \text{subject to} && \mathbf{x} \in \mathcal{X}, \omega \in \{0, 1\}^S, \mathbf{y} \in \mathcal{Y} \\
& && \sum_{s \in \mathcal{S}} \omega^s = 1, \quad \omega^s = 1 \Leftrightarrow \mathbf{y} = \bar{\mathbf{y}}^s \quad \forall s \in \mathcal{S} \\
& && \mathbf{A}^s(\boldsymbol{\xi}^s) \mathbf{x} + \mathbf{D}^s(\boldsymbol{\xi}^s) \mathbf{y} + \tilde{\mathbf{A}}^s(\boldsymbol{\xi}^s) \tilde{\mathbf{x}}(\mathbf{y} \circ \boldsymbol{\xi}) \leq \mathbf{b}^s(\boldsymbol{\xi}^s) + M^s \mathbf{e}(1 - \omega^s) \quad \forall s \in \mathcal{S}, \boldsymbol{\xi} \in \Xi(\mathbf{x}, \omega, \mathbf{y}) \\
& && \tilde{\mathbf{x}}(\mathbf{y} \circ \boldsymbol{\xi}) \in \tilde{\mathcal{X}} \quad \forall \boldsymbol{\xi} \in \Xi(\mathbf{x}, \omega, \mathbf{y})
\end{aligned} \tag{9}$$

with

$$\Xi(\mathbf{x}, \omega, \mathbf{y}) = \left\{ \boldsymbol{\xi} \in \mathbb{R}^K : \begin{array}{l} \mathbf{W}^s \boldsymbol{\xi}^s \leq \mathbf{U}^s \mathbf{x} + \widehat{M}^s \mathbf{e}(1 - \omega^s) \quad \forall s \in \mathcal{S} \\ \boldsymbol{\xi}^{\min} \circ \mathbf{y} \leq \boldsymbol{\xi} \leq \boldsymbol{\xi}^{\max} \circ \mathbf{y} \end{array} \right\},$$

where M^s and \widehat{M}^s are vectors of sufficiently large parameters. As the statement $\omega^s = 1 \Leftrightarrow \mathbf{y} = \bar{\mathbf{y}}^s$ can be expressed as a set of linear constraints, e.g. $\bar{\mathbf{y}}^s \omega^s \leq \mathbf{y} \leq \mathbf{e} + (\bar{\mathbf{y}}^s - \mathbf{e})\omega^s$, formulation (9) can be easily transformed into the form of (5) by appropriate redefinition of variables, constraint matrices, and right-hand sides. \square

4.3 Incorporating Type-2b Endogenous Uncertainty

Finally, we introduce additional binary variables $z \in \{0, 1\}^K$ to indicate which uncertain parameters are observed, and we arrive at the following unified two-stage robust optimization formulation that considers type-1, type-2a, and type-2b endogenous uncertainty:

$$\begin{aligned}
& \underset{\mathbf{x}, \mathbf{y}, \mathbf{z}, \tilde{\mathbf{x}}}{\text{minimize}} && x_1 \\
& \text{subject to} && \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}, \mathbf{z} \in \mathcal{Z}, \mathbf{z} \leq \mathbf{y} \\
& && \left. \begin{array}{l} \mathbf{A}(\boldsymbol{\xi}) \mathbf{x} + \mathbf{D}(\boldsymbol{\xi}) \mathbf{y} + \mathbf{H}(\boldsymbol{\xi}) \mathbf{z} + \tilde{\mathbf{A}}(\boldsymbol{\xi}) \tilde{\mathbf{x}}(\mathbf{z} \circ \boldsymbol{\xi}) \leq \mathbf{b}(\boldsymbol{\xi}) \\ \tilde{\mathbf{x}}(\mathbf{z} \circ \boldsymbol{\xi}) \in \tilde{\mathcal{X}} \end{array} \right\} \quad \forall \boldsymbol{\xi} \in \Xi(\mathbf{x}, \mathbf{y}),
\end{aligned} \tag{10}$$

where $\mathcal{Z} \subseteq \{0, 1\}^K$. The recourse decisions $\tilde{\mathbf{x}}(\mathbf{z} \circ \boldsymbol{\xi})$ can only be functions of the observed parameters. Also, by requiring $\mathbf{z} \leq \mathbf{y}$, parameters can only be observed if they materialize. The formulation of the uncertainty set does not change from (6):

$$\Xi(\mathbf{x}, \mathbf{y}) = \{ \boldsymbol{\xi} \in \mathbb{R}^K : \mathbf{W} \boldsymbol{\xi} \leq \mathbf{U} \mathbf{x} + \bar{\mathbf{U}} \mathbf{y} \}.$$

5 Multistage Robust Optimization with Endogenous Uncertainty

In multistage robust optimization with endogenous uncertainty, decisions that affect the uncertainty can also be recourse variables. Moreover, while the two-stage problem can only consider decisions determining *if* uncertain parameters materialize or are observed, multistage formulations can incorporate decisions affecting *when*, i.e. in which stage, uncertain parameters material-

ize or are observed.

Let $\mathcal{T} = \{1, \dots, T\}$ denote the set of stages and \mathcal{I} the set of uncertain parameters. An uncertain parameter $i \in \mathcal{I}$ can materialize anytime between (and including) given stages τ_i and $\bar{\tau}_i$. We define a vector of uncertain parameters $\xi_t \in \mathbb{R}^{K_t}$ that contains all uncertain parameters i for which $\tau_i = t$. We further define all uncertain parameters that can materialize in or prior to stage t as $\xi_{[t]} = (\xi_1, \dots, \xi_t)$, and $\xi = \xi_{[T]}$. The set of uncertain parameters for which $\tau_i = t$ is denoted by $\tilde{\mathcal{I}}_t$, and we have $\tilde{\mathcal{I}}_t \cap \tilde{\mathcal{I}}_{t'} = \emptyset$ for all $t \neq t'$ and $\mathcal{I} = \bigcup_{t \in \mathcal{T}} \tilde{\mathcal{I}}_t$. The set of uncertain parameters that can materialize in stage t is denoted by $\bar{\mathcal{I}}_t$. In the special case where every uncertain parameter can only materialize in one stage, $\bar{\mathcal{I}}_t = \tilde{\mathcal{I}}_t$. Similarly, let $\hat{\mathcal{I}}_t$ denote the set of uncertain parameters that can be observed in stage t , and $\hat{\tau}_i$ and $\tilde{\tau}_i$ be the stages between which uncertain parameter i can be observed. Since the observation of an uncertain parameter can only occur after its materialization, we have $\hat{\tau}_i \geq \tau_i$ and $\tilde{\tau}_i \geq \bar{\tau}_i$.

The general multistage robust optimization problem with type-1, type-2a, and type-2b endogenous uncertainty can be formulated as follows:

$$\begin{aligned} & \underset{\mathbf{x}, \mathbf{y}, \mathbf{z}}{\text{minimize}} && x_{11} \\ & \text{subject to} && \left. \begin{aligned} & \sum_{t'=1}^t [A_{tt'}(\xi_{[t]})\mathbf{x}_{t'} + D_{tt'}(\xi_{[t]})\mathbf{y}_{t'} + H_{tt'}(\xi_{[t]})\mathbf{z}_{t'}] \leq \mathbf{b}_t(\xi_{[t]}) \\ & \mathbf{x}_t = \mathbf{x}_t(\bar{\mathbf{z}}_{t-1} \circ \xi_{[t]}), \mathbf{x}_t \in \mathcal{X}_t \\ & \mathbf{y}_t = \mathbf{y}_t(\bar{\mathbf{z}}_{t-1} \circ \xi_{[t]}), \mathbf{y}_t \in \mathcal{Y}_t \\ & \mathbf{z}_t = \mathbf{z}_t(\bar{\mathbf{z}}_{t-1} \circ \xi_{[t]}), \mathbf{z}_t \in \mathcal{Z}_t \end{aligned} \right\} \quad (11) \\ & && \forall t \in \mathcal{T}, \xi_{[t]} \in \Xi_t(\mathbf{x}_{[t-1]}, \mathbf{y}_{[t-1]}, \mathbf{z}_{[t-2]}) \end{aligned}$$

where $\mathcal{Y}_t \in \{0, 1\}^{|\tilde{\mathcal{I}}_{t+1}|}$ and $\mathcal{Z}_t \in \{0, 1\}^{|\hat{\mathcal{I}}_{t+1}|}$. The notation is such that $\mathbf{x}_{[t]} = (\mathbf{x}_1, \dots, \mathbf{x}_t)$, $\mathbf{x} = \mathbf{x}_{[T]}$, $\mathbf{y}_{[t]} = (\mathbf{y}_1, \dots, \mathbf{y}_t)$, $\mathbf{y} = \mathbf{y}_{[T]}$, $\mathbf{z}_{[t]} = (\mathbf{z}_1, \dots, \mathbf{z}_t)$, and $\mathbf{z} = \mathbf{z}_{[T]}$. The binary variable y_{ti} equals 1 if uncertain parameter i materializes in stage $t + 1$, whereas the binary variable z_{ti} equals 1 if uncertain parameter i is observed in stage $t + 1$. We use auxiliary variables $\bar{\mathbf{z}}_{t-1} = \left(\sum_{t'=0}^{t-1} z_{t'1}, \dots, \sum_{t'=0}^{t-1} z_{t'K_{[t]}} \right)$ with $K_{[t]} = \sum_{t'=1}^t K_{t'}$ to indicate which uncertain parameters in $\xi_{[t]}$ are observed up to stage t . The recourse variables in stage t , i.e. \mathbf{x}_t , \mathbf{y}_t , and \mathbf{z}_t , are functions of $\bar{\mathbf{z}}_{t-1} \circ \xi_{[t]}$, which represent the uncertain parameters observed up to stage t . For notational convenience, we assume that $\xi_1 = \xi_{11} = 1$, $\mathbf{y}_0 = \mathbf{y}_{01} = 1$, and $\mathbf{z}_0 = \mathbf{z}_{01} = 1$. The stage-specific uncertainty set Ξ_t is defined such that $\Xi_1 = \{\xi_1 \in \mathbb{R} : \xi_{11} = 1\}$, and for $t \geq 2$:

$$\Xi_t(\mathbf{x}_{[t-1]}, \mathbf{y}_{[t-1]}, \mathbf{z}_{[t-2]}) = \left\{ \xi_{[t]} \in \mathbb{R}^{K_{[t]}} : \begin{aligned} & \mathbf{W}_t \xi_{[t]} \leq \sum_{t'=1}^{t-1} [\mathbf{U}_{tt'} \mathbf{x}_{t'} + \bar{\mathbf{U}}_{tt'} \mathbf{y}_{t'}] \\ & \mathbf{x}_{t'} = \mathbf{x}_{t'}(\bar{\mathbf{z}}_{t-1} \circ \xi_{[t']}) \quad \forall t' = 1, \dots, t-1 \\ & \mathbf{y}_{t'} = \mathbf{y}_{t'}(\bar{\mathbf{z}}_{t-1} \circ \xi_{[t']}) \quad \forall t' = 1, \dots, t-1 \\ & \mathbf{z}_{t'} = \mathbf{z}_{t'}(\bar{\mathbf{z}}_{t-1} \circ \xi_{[t']}) \quad \forall t' = 1, \dots, t-1 \end{aligned} \right\}, \quad (12)$$

where the polyhedral set directly depends on $\mathbf{x}_{[t-1]}$ and $\mathbf{y}_{[t-1]}$, which in turn are functions of observed uncertain parameters given by $\mathbf{z}_{[t-2]}$. We assume that $\Xi_t(\mathbf{x}_{[t-1]}, \mathbf{y}_{[t-1]}, \mathbf{z}_{[t-2]}) \subseteq \{\xi_{[t]} \in \mathbb{R}^{K_{[t]}} : \xi_{[t]}^{\min} \leq \xi_{[t]} \leq \xi_{[t]}^{\max}\}$ for all feasible $(\mathbf{x}_{[t-1]}, \mathbf{y}_{[t-1]}, \mathbf{z}_{[t-2]})$.

The linear constraints in problem (11) include the following inequalities:

$$\sum_{t'=\tau_i}^{\bar{\tau}_i} y_{t'-1,i} \leq 1 \quad \forall i \in \mathcal{I} \quad (13a)$$

$$\sum_{t'=\hat{\tau}_i}^{\bar{\tau}_i} z_{t'-1,i} \leq 1 \quad \forall i \in \mathcal{I} \quad (13b)$$

$$\sum_{t'=\tau_i}^{\min\{t, \bar{\tau}_i\}} y_{t'-1,i} \geq z_{t-1,i} \quad \forall t \in \mathcal{T} \setminus \{1\}, i \in \hat{\mathcal{I}}_t. \quad (13c)$$

Constraints (13a) state that each uncertain parameter i can only materialize between stages τ_i and $\bar{\tau}_i$; analogously, uncertain parameter i can only be observed between stages $\hat{\tau}_i$ and $\bar{\tau}_i$ according to (13b). Inequalities (13c) ensure that an uncertain parameter can only be observed after it has materialized.

6 Decision Rule Approach

In this section, we present a decision rule approach to (approximately) solve the general multistage robust optimization problem under endogenous uncertainty, which reduces to the two-stage case when $T = 2$. The proposed approach is an extension of the method presented in Feng et al.⁴⁸ and relies on the concept of lifted uncertainty⁵⁵. Here, the main innovation is the use of auxiliary uncertain parameters to model decision-dependent nonanticipativity.

For ease of exposition and for tractability reasons, we consider the case of fixed recourse, i.e. $A_{tt'}$, $D_{tt'}$, and $H_{tt'}$ do not depend on $\xi_{[t]}$. In theory, the presented approach can be adapted to consider random recourse¹¹, however, at considerable additional computational cost. We further assume that $b_t(\xi_{[t]}) = B_t \xi_{[t]}$ and $\mathcal{X}_t = \mathbb{R}^{\bar{N}_t} \times \{0, 1\}^{\hat{N}_t}$. Note that x_t contain both continuous and binary variables, and we define $x_t = (x_t^c, x_t^b)$ such that $x_t^c \in \mathbb{R}^{\bar{N}_t}$ and $x_t^b \in \{0, 1\}^{\hat{N}_t}$.

6.1 Decision-Dependent Nonanticipativity

In problem (11), recourse variables can only be adjusted based on the values of observed uncertain parameters; this is referred to as nonanticipativity. As a result, any decision rule applied to a recourse variable can only be a function of uncertain parameters that have been observed up to the corresponding stage. Suppose we apply decision rules with a linear structure such that, for example, $x_t = \tilde{X}_t \xi_{[t]}$. Then, intuitively, one may attempt to model nonanticipativity by incorporating additional constraints that force decision rule parameters corresponding to unobserved uncertain parameters to be zero, e.g. $-M \bar{z}_{t-1,i} e \leq \tilde{x}_{ti} \leq M \bar{z}_{t-1,i} e$ with M being a sufficiently large parameter. In the two-stage case, where z are only first-stage variables, this is a viable approach. However, in a multistage setting, $z_{[t-1]}$ can themselves be adjustable recourse variables; in that case, the added constraints also require \tilde{X}_t to be adjustable. Simply applying another decision rule to \tilde{X}_t does not resolve the issue since the decision rule parameters for \tilde{X}_t would also have to be adjustable because of their dependence on $z_{[t-1]}$. For fixed uncertainty sets, Vayanos et al.³⁸ propose an approach that divides the uncertainty set into a number of preselected subsets

\mathcal{S} and introduces recourse variables for each subset $s \in \mathcal{S}$. As a result, constant z_s can be selected for each subset s , which allows restrictions to be imposed on the decision rule parameters for the remaining recourse variables such that nonanticipativity is satisfied. However, this approach is severely computationally intractable if the number of uncertainty subsets is large, which is only exacerbated in the case of problem (11) where the uncertainty set is generally not fixed.

We propose an alternative, more tractable approach to incorporate decision-dependent nonanticipativity. Notice that given a linear decision rule structure, the dependence of a recourse variable on an uncertain parameter is equally eliminated if, instead of setting the corresponding decision rule coefficients to zero, the uncertain parameter itself is set to zero. In the special case where an uncertain parameter can only materialize in one stage and its materialization directly leads to its observation (i.e. $\mathbf{y} = \mathbf{z}$), this can be easily incorporated by including the following inequalities in the definition of the uncertainty set Ξ_t :

$$\xi_{t'}^{\min} \circ \mathbf{y}_{t'-1} \leq \xi_{t'} \leq \xi_{t'}^{\max} \circ \mathbf{y}_{t'-1} \quad \forall t' = 1, \dots, t, \quad (14)$$

which are valid since, as noted in the proof of Proposition 1, the values of unmaterialized parameters can be set to zero without loss of generality. However, inequalities (14) cannot capture the general case where there may be multiple stages in which an uncertain parameter can materialize or where despite being already materialized, one can still decide whether or not to observe the true value of the uncertain parameter in a given stage.

To address the general case, we first generalize (14) such that uncertain parameters that can materialize in multiple stages are considered:

$$\xi_{\tau_i i}^{\min} \sum_{t'=\tau_i}^{\min\{t, \bar{\tau}_i\}} y_{t'-1, i} \leq \xi_{\tau_i i} \leq \xi_{\tau_i i}^{\max} \sum_{t'=\tau_i}^{\min\{t, \bar{\tau}_i\}} y_{t'-1, i} \quad \forall i \in \bigcup_{t'=1}^t \tilde{\mathcal{I}}_{t'}. \quad (15)$$

We then introduce $\bar{\tau}_i - \hat{\tau}_i + 1$ auxiliary uncertain parameters $\zeta_{\hat{\tau}_i i}, \dots, \zeta_{\bar{\tau}_i i}$ for every $i \in \mathcal{I}$. Note that for an original uncertain parameter i for which $\tau_i = \bar{\tau}_i$ and $y_{\tau_i-1, i} = z_{\tau_i-1, i}$ (which corresponds to the above-mentioned special case), we do not actually require auxiliary uncertain parameters; however, for ease of notation, we assume that we have auxiliary uncertain parameters associated with every $i \in \mathcal{I}$. We incorporate the following inequalities in the definition of the uncertainty set for stage t :

$$\xi_{\tau_i i} - \xi_{\tau_i i}^{\max} (1 - z_{t'-1, i}) \leq \zeta_{t' i} \leq \xi_{\tau_i i} + M(1 - z_{t'-1, i}) \quad \forall t' = 1, \dots, t, i \in \hat{\mathcal{I}}_{t'} \quad (16a)$$

$$\xi_{\tau_i i}^{\min} z_{t'-1, i} \leq \zeta_{t' i} \leq \xi_{\tau_i i}^{\max} z_{t'-1, i} \quad \forall t' = 1, \dots, t, i \in \tilde{\mathcal{I}}_{t'}, \quad (16b)$$

where M is a sufficiently large big-M parameter. Inequalities (16) enforce that $\zeta_{t' i}$ takes the value of $\xi_{\tau_i i}$ if $\xi_{\tau_i i}$ is observed in stage t' , and is zero otherwise.

With a slight abuse of notation, we define the following new uncertainty set that considers the

original and the auxiliary uncertain parameters:

$$\begin{aligned} \Theta_t(\mathbf{x}_{[t-1]}, \mathbf{y}_{[t-1]}, \mathbf{z}_{[t-1]}) = & \left\{ (\boldsymbol{\xi}_{[t]}, \boldsymbol{\zeta}_{[t]}) \in \mathbb{R}^{\sum_{t'=1}^t K_{t'} + |\widehat{\mathcal{I}}_{t'}|} : \sum_{t'=1}^t [\mathbf{W}_{tt'} \boldsymbol{\xi}_{t'} + \overline{\mathbf{W}}_{tt'} \boldsymbol{\zeta}_{t'}] \right. \\ & \left. \leq \sum_{t'=1}^{t-1} \left[\mathbf{U}_{tt'}^c \mathbf{x}_{t'}^c(\boldsymbol{\zeta}_{[t']}) + \mathbf{U}_{tt'}^b \mathbf{x}_{t'}^b(\boldsymbol{\zeta}_{[t']}) + \overline{\mathbf{U}}_{tt'} \mathbf{y}_{t'}(\boldsymbol{\zeta}_{[t']}) + \widehat{\mathbf{U}}_{tt'} \mathbf{z}_{t'}(\boldsymbol{\zeta}_{[t']}) \right] \right\}, \end{aligned}$$

where $\mathbf{W}_{tt'} \in \mathbb{R}^{Q_t \times K_{t'}}$, $\overline{\mathbf{W}}_{tt'} \in \mathbb{R}^{Q_t \times |\widehat{\mathcal{I}}_{t'}|}$, $\mathbf{U}_{tt'}^c \in \mathbb{R}^{Q_t \times \bar{N}_{t'}}$, $\mathbf{U}_{tt'}^b \in \mathbb{R}^{Q_t \times \widehat{N}_{t'}}$, $\overline{\mathbf{U}}_{tt'} \in \mathbb{R}^{Q_t \times |\bar{\mathcal{I}}_{t'+1}|}$, and $\widehat{\mathbf{U}}_{tt'} \in \mathbb{R}^{Q_t \times |\widehat{\mathcal{I}}_{t'+1}|}$ are chosen to incorporate the inequalities from the original uncertainty set (12) as well as inequalities (15) and (16). Notice that the recourse variables \mathbf{x}_t^c , \mathbf{x}_t^b , \mathbf{y}_t , and \mathbf{z}_t now depend on $\boldsymbol{\zeta}_{[t]}$, which represent the uncertain parameters observed up to stage t .

6.2 Lifted Uncertainty

For each auxiliary uncertain parameter ζ_{ti} , we introduce two sets of new uncertain parameters and hence create a higher-dimensional *lifted* uncertainty space, which will facilitate the design of flexible decision rules. To this end, the bounded marginal support of ζ_{ti} , which is the same as the one of $\xi_{\tau_i i}$, is partitioned into r_i pieces defined by its lower and upper bounds as well as $r_i - 1$ breakpoints, and we denote these points by p_i^j with $j = 0, 1, \dots, r_i$ such that

$$\xi_{\tau_i i}^{\min} = p_i^0 < p_i^1 < \dots < p_i^{r_i-1} < p_i^{r_i} = \xi_{\tau_i i}^{\max}.$$

We then define two lifting operators, $\bar{L}_t : \mathbb{R}^{|\widehat{\mathcal{I}}_t|} \mapsto \mathbb{R}^{\bar{K}_t}$ and $\widehat{L}_t : \mathbb{R}^{|\widehat{\mathcal{I}}_t|} \mapsto \{0, 1\}^{\widehat{K}_t}$, for each $t \in \mathcal{T}$. The first lifting operator $\bar{L}_t(\boldsymbol{\zeta}_t)$ maps $\boldsymbol{\zeta}_t$ onto a \bar{K}_t -dimensional space with $\bar{K}_t = \sum_{i \in \widehat{\mathcal{I}}_t} r_i$ such that each lifted parameter is a piecewise linear function of the corresponding ζ_{ti} and is defined as follows:

$$\bar{L}_{ti}^j(\boldsymbol{\zeta}_t) := \begin{cases} \zeta_{ti} & \text{if } r_i = 1 \\ \inf \{ \zeta_{ti}, p_i^j \} & \text{if } r_i > 1, j = 1 \\ \sup \{ \inf \{ \zeta_{ti}, p_i^j \} - p_i^{j-1}, 0 \} & \text{if } r_i > 1, j = 2, \dots, r_i. \end{cases}$$

By construction, we have $\zeta_{ti} = \sum_{j=1}^{r_i} \bar{L}_{ti}^j(\boldsymbol{\zeta}_t)$. The second lifting operator $\widehat{L}_t(\boldsymbol{\zeta}_t)$ maps $\boldsymbol{\zeta}_t$ onto a \widehat{K}_t -dimensional space with $\widehat{K}_t = \sum_{i \in \widehat{\mathcal{I}}_t} g_i$ and $g_i = \max\{1, r_i - 1\}$. It is defined such that

$$\widehat{L}_{ti}^j(\boldsymbol{\zeta}_t) := \begin{cases} 1 & \text{if } r_i = 1 \\ \mathbb{1}(\zeta_{ti} \geq p_i^j) & \text{if } r_i > 1, j = 1, \dots, g_i, \end{cases}$$

which is a piecewise constant function of ζ_{ti} .

Using the lifting operators, we can define the following lifted endogenous uncertainty set:

$$\tilde{\Theta}'_t(\mathbf{x}_{[t-1]}, \mathbf{y}_{[t-1]}, \mathbf{z}_{[t-1]}) = \left\{ (\boldsymbol{\xi}_{[t]}, \boldsymbol{\zeta}_{[t]}, \bar{\boldsymbol{\zeta}}_{[t]}, \hat{\boldsymbol{\zeta}}_{[t]}) \in \mathbb{R}^{\tilde{K}_{[t]}} : \right. \\ \left. \sum_{t'=1}^t [\mathbf{W}_{tt'} \boldsymbol{\xi}_{t'} + \bar{\mathbf{W}}_{tt'} \boldsymbol{\zeta}_{t'}] \leq \sum_{t'=1}^{t-1} \left[\mathbf{U}_{tt'}^c \mathbf{x}_{t'}^c(\boldsymbol{\zeta}_{[t']}) + \mathbf{U}_{tt'}^b \mathbf{x}_{t'}^b(\boldsymbol{\zeta}_{[t']}) + \bar{\mathbf{U}}_{tt'} \mathbf{y}_{t'}(\boldsymbol{\zeta}_{[t']}) + \hat{\mathbf{U}}_{tt'} \mathbf{z}_{t'}(\boldsymbol{\zeta}_{[t']}) \right] \right. \\ \left. \bar{\boldsymbol{\zeta}}_{t'} = \bar{\mathbf{L}}_{t'}(\boldsymbol{\zeta}_{t'}), \hat{\boldsymbol{\zeta}}_{t'} = \hat{\mathbf{L}}_{t'}(\boldsymbol{\zeta}_{t'}) \quad \forall t' = 1, \dots, t \right\},$$

where $\tilde{K}_{[t]} = \sum_{t'=1}^t K_{t'} + |\hat{\mathcal{I}}_{t'}| + \bar{K}_{t'} + \hat{K}_{t'}$, and for all $i \in \hat{\mathcal{I}}_t$, $\bar{\boldsymbol{\zeta}}_{ti} = (\bar{\zeta}_{ti}^1, \dots, \bar{\zeta}_{ti}^{r_i})$, $\bar{\mathbf{L}}_{ti} = (\bar{L}_{ti}^1, \dots, \bar{L}_{ti}^{r_i})$, $\hat{\boldsymbol{\zeta}}_{ti} = (\hat{\zeta}_{ti}^1, \dots, \hat{\zeta}_{ti}^{g_i})$, and $\hat{\mathbf{L}}_{ti} = (\hat{L}_{ti}^1, \dots, \hat{L}_{ti}^{g_i})$. The main purpose of defining such a lifted uncertainty set is to enable the construction of linear decision rules in the space of the lifted uncertain parameters. Notice that $\tilde{\Theta}'_t$ is an open set due to the discontinuity in $\hat{\mathbf{L}}$. Now consider problem (11) with Ξ_t replaced by $\tilde{\Theta}'_t$ and the recourse variables restricted to be linear functions of $\bar{\boldsymbol{\zeta}}$ and $\hat{\boldsymbol{\zeta}}$, and a second problem that is the same except that it considers the convex hull of the closure of $\tilde{\Theta}'_t$, i.e. $\text{conv}(\text{cl}(\tilde{\Theta}'_t))$, instead of $\tilde{\Theta}'_t$. Following the same arguments as in Bertsimas and Georghiou¹¹, one can show that these two problems are equivalent in a sense that they have the same optimal value and that there is a one-to-one mapping between their feasible and optimal solutions.

It is difficult to obtain a closed-form description of $\text{conv}(\text{cl}(\tilde{\Theta}'_t))$ as it requires finding all vertices of $\text{cl}(\tilde{\Theta}'_t)$, which are decision-dependent. Moreover, the number of vertices grows exponentially with the dimensionality of $\text{cl}(\tilde{\Theta}'_t)$ such that including them would render most problems of practical relevance computationally intractable. Therefore, instead of $\text{conv}(\text{cl}(\tilde{\Theta}'_t))$, we consider an outer approximation, which we construct as follows. For each $t \in \mathcal{T}$, $i \in \hat{\mathcal{I}}_t$, and $j = 1, \dots, r_i$, we define a set of vertices $\tilde{\mathcal{V}}_{ti}^j$ in the lifted space:

$$\tilde{\mathcal{V}}_{ti}^j := \left\{ \tilde{\mathbf{v}} = (v, \bar{\mathbf{v}}, \hat{\mathbf{v}}) : \begin{array}{l} v \in \{p_i^{j-1}, p_i^j\} \\ \bar{\mathbf{v}} = \bar{\mathbf{L}}_{ti}(\mathbf{e}_{ti}v) \\ \hat{\mathbf{v}} = \lim_{\zeta_{ti} \rightarrow v, \zeta_{ti} \in [p_i^{j-1}, p_i^j]} \hat{\mathbf{L}}_{ti}(\mathbf{e}_{ti}\zeta_{ti}) \end{array} \right\},$$

which allows the following convex hull formulation of the closure of the marginal support of $\tilde{\boldsymbol{\zeta}}_{ti}$:

$$Z_{ti} = \left\{ \tilde{\boldsymbol{\zeta}}_{ti} = (\zeta_{ti}, \bar{\boldsymbol{\zeta}}_{ti}, \hat{\boldsymbol{\zeta}}_{ti}) : \begin{array}{l} \sum_{\tilde{\mathbf{v}} \in \tilde{\mathcal{V}}_{ti}} \lambda(\tilde{\mathbf{v}}) = 1 \\ \zeta_{ti} = \sum_{\tilde{\mathbf{v}} \in \tilde{\mathcal{V}}_{ti}} \lambda(\tilde{\mathbf{v}})v \\ \bar{\boldsymbol{\zeta}}_{ti} = \sum_{\tilde{\mathbf{v}} \in \tilde{\mathcal{V}}_{ti}} \lambda(\tilde{\mathbf{v}})\bar{\mathbf{v}} \\ \hat{\boldsymbol{\zeta}}_{ti} = \sum_{\tilde{\mathbf{v}} \in \tilde{\mathcal{V}}_{ti}} \lambda(\tilde{\mathbf{v}})\hat{\mathbf{v}} \\ \lambda(\tilde{\mathbf{v}}) \in \mathbb{R}_+ \quad \forall \tilde{\mathbf{v}} \in \tilde{\mathcal{V}}_{ti} \end{array} \right\},$$

where $\tilde{\mathcal{V}}_{ti} = \bigcup_{j=1}^{r_i} \tilde{\mathcal{V}}_{ti}^j$ and $\lambda(\tilde{\mathbf{v}})$ denotes the coefficient associated with a particular vertex $\tilde{\mathbf{v}} \in \tilde{\mathcal{V}}_{ti}$. We can now obtain the following closed-form expression of an outer approximation of

$\text{conv}(\text{cl}(\tilde{\Xi}_t')):$

$$\left\{ \{\xi_{[t]}\} \times \prod_{t'=1}^t \prod_{i \in \hat{\mathcal{I}}_{t'}} Z_{t'i} \right\} \cap \left\{ (\xi_{[t]}, \zeta_{[t]}, \bar{\zeta}_{[t]}, \hat{\zeta}_{[t]}) \in \mathbb{R}^{\tilde{K}_{[t]}} : \sum_{t'=1}^t [\mathbf{w}_{tt'} \xi_{t'} + \bar{\mathbf{w}}_{tt'} \zeta_{t'}] \right. \\ \left. \leq \sum_{t'=1}^{t-1} [U_{tt'}^c \mathbf{x}_{t'}^c(\zeta_{[t']}) + U_{tt'}^b \mathbf{x}_{t'}^b(\zeta_{[t']}) + \bar{U}_{tt'} \mathbf{y}_{t'}(\zeta_{[t']}) + \hat{U}_{tt'} \mathbf{z}_{t'}(\zeta_{[t']})] \right\}. \quad (17)$$

6.3 Decision Rule Approximation

Following a similar approach as in Feng et al. (2020), we apply decision rules that are linear functions of the lifted uncertain parameters and parameterized as follows:

$$\mathbf{x}_t^c = \sum_{t'=1}^t \sum_{i \in \hat{\mathcal{I}}_{t'}} \left(\bar{\mathbf{X}}_{tt'i}^c \bar{\zeta}_{t'i} + \hat{\mathbf{X}}_{tt'i}^c \hat{\zeta}_{t'i} \right) \quad (18a)$$

$$\mathbf{x}_t^b = \sum_{t'=1}^t \sum_{i \in \hat{\mathcal{I}}_{t'}} \hat{\mathbf{X}}_{tt'i}^b \hat{\zeta}_{t'i}, \quad \mathbf{y}_t = \sum_{t'=1}^t \sum_{i \in \hat{\mathcal{I}}_{t'}} \hat{\mathbf{Y}}_{tt'i} \hat{\zeta}_{t'i}, \quad \mathbf{z}_t = \sum_{t'=1}^t \sum_{i \in \hat{\mathcal{I}}_{t'}} \hat{\mathbf{Z}}_{tt'i} \hat{\zeta}_{t'i}, \quad (18b)$$

where $\bar{\mathbf{X}}_{tt'i}^c \in \mathbb{R}^{\bar{N}_t \times r_i}$, $\hat{\mathbf{X}}_{tt'i}^c \in \mathbb{R}^{\bar{N}_t \times g_i}$, $\hat{\mathbf{X}}_{tt'i}^b \in \{-1, 0, 1\}^{\bar{N}_t \times g_i}$, $\hat{\mathbf{Y}}_{tt'i} \in \{-1, 0, 1\}^{|\bar{\mathcal{I}}_{t+1}| \times g_i}$, and $\hat{\mathbf{Z}}_{tt'i} \in \{-1, 0, 1\}^{|\bar{\mathcal{I}}_{t+1}| \times g_i}$. The decision rules for the continuous variables in (18a) can form continuous or discontinuous piecewise linear functions of $\zeta_{[t]}$. With the given domains for $\hat{\mathbf{X}}_{tt'i}^b$, $\hat{\mathbf{Y}}_{tt'i}$, and $\hat{\mathbf{Z}}_{tt'i}$, a variable following a decision rule of the form (18b) is guaranteed to be binary, i.e. can only take the value 0 or 1, if we also constrain it to be in $[0, 1]$ ¹¹. This property implies that the integrality constraints on the binary variables can be relaxed when these decision rules are applied.

By applying the proposed decision rules in conjunction with the uncertainty set in (17), decision-dependent nonanticipativity is captured, and we arrive at the following approximation

of the multistage robust optimization problem under endogenous uncertainty:

$$\begin{aligned}
& \underset{\substack{\overline{\mathbf{X}}^c, \widehat{\mathbf{X}}^c, \\ \widehat{\mathbf{X}}^b, \widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}}}{\text{minimize}} & \mathbf{e}_1^\top \left(\overline{\mathbf{X}}_{111}^c + \widehat{\mathbf{X}}_{111}^c \right) \\
& \text{subject to} & \left(\overline{\mathbf{X}}_{tt'i}^c, \widehat{\mathbf{X}}_{tt'i}^c, \widehat{\mathbf{X}}_{tt'i}^b, \widehat{\mathbf{Y}}_{tt'i}, \widehat{\mathbf{Z}}_{tt'i} \right) \in \Omega_{tt'i} \quad \forall t \in \mathcal{T}, t' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t'} \\
& & \left. \begin{aligned} & \sum_{t'=1}^t \sum_{t''=1}^{t'} \sum_{i \in \widehat{\mathcal{I}}_{t''}} \left[\mathbf{A}_{tt'}^c \left(\overline{\mathbf{X}}_{t't''i}^c \bar{\boldsymbol{\zeta}}_{t''i} + \widehat{\mathbf{X}}_{t't''i}^c \hat{\boldsymbol{\zeta}}_{t''i} \right) + \mathbf{A}_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b \hat{\boldsymbol{\zeta}}_{t''i} \right. \\ & \quad \left. + \mathbf{D}_{tt'} \widehat{\mathbf{Y}}_{t't''i} \hat{\boldsymbol{\zeta}}_{t''i} + \mathbf{H}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \hat{\boldsymbol{\zeta}}_{t''i} \right] \leq \mathbf{B}_t \boldsymbol{\xi}_{[t]} \\ & \mathbf{0} \leq \sum_{t''=1}^{t'} \sum_{i \in \widehat{\mathcal{I}}_{t''}} \widehat{\mathbf{X}}_{t't''i}^b \hat{\boldsymbol{\zeta}}_{t''i} \leq \mathbf{e} \quad \forall t' = 1, \dots, t \\ & \mathbf{0} \leq \sum_{t''=1}^{t'} \sum_{i \in \widehat{\mathcal{I}}_{t''}} \widehat{\mathbf{Y}}_{t't''i} \hat{\boldsymbol{\zeta}}_{t''i} \leq \mathbf{e} \quad \forall t' = 1, \dots, t \\ & \mathbf{0} \leq \sum_{t''=1}^{t'} \sum_{i \in \widehat{\mathcal{I}}_{t''}} \widehat{\mathbf{Z}}_{t't''i} \hat{\boldsymbol{\zeta}}_{t''i} \leq \mathbf{e} \quad \forall t' = 1, \dots, t \end{aligned} \right\} \quad (19) \\
& & \forall t \in \mathcal{T}, (\boldsymbol{\xi}_{[t]}, \boldsymbol{\zeta}_{[t]}, \bar{\boldsymbol{\zeta}}_{[t]}, \hat{\boldsymbol{\zeta}}_{[t]}) \in \tilde{\Theta}_t \left(\overline{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^b, \widehat{\mathbf{Y}}_{[t-1]}, \widehat{\mathbf{Z}}_{[t-1]} \right),
\end{aligned}$$

where

$$\Omega_{tt'i} = \left\{ \left(\overline{\mathbf{X}}_{tt'i}^c, \widehat{\mathbf{X}}_{tt'i}^c, \widehat{\mathbf{X}}_{tt'i}^b, \widehat{\mathbf{Y}}_{tt'i}, \widehat{\mathbf{Z}}_{tt'i} \right) : \begin{aligned} & \overline{\mathbf{X}}_{tt'i}^c \in \mathbb{R}^{\bar{N}_t \times r_i} \\ & \widehat{\mathbf{X}}_{tt'i}^c \in \mathbb{R}^{\bar{N}_t \times g_i} \\ & \widehat{\mathbf{X}}_{tt'i}^b \in \{-1, 0, 1\}^{\bar{N}_t \times g_i} \\ & \widehat{\mathbf{Y}}_{tt'i} \in \{-1, 0, 1\}^{|\bar{\mathcal{I}}_{t+1}| \times g_i} \\ & \widehat{\mathbf{Z}}_{tt'i} \in \{-1, 0, 1\}^{|\widehat{\mathcal{I}}_{t+1}| \times g_i} \end{aligned} \right\},$$

and $\mathbf{A}_{tt'}^c$ and $\mathbf{A}_{tt'}^b$ are appropriately defined constraint matrices associated with the corresponding continuous and binary \mathbf{x} -variables, respectively. The uncertainty set is

$$\begin{aligned}
& \tilde{\Theta}_t \left(\overline{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^b, \widehat{\mathbf{Y}}_{[t-1]}, \widehat{\mathbf{Z}}_{[t-1]} \right) \\
& = \left\{ \{\boldsymbol{\xi}_{[t]}\} \times \prod_{t'=1}^t \prod_{i \in \widehat{\mathcal{I}}_{t'}} Z_{t'i} \right\} \cap \hat{\Theta}_t \left(\overline{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^b, \widehat{\mathbf{Y}}_{[t-1]}, \widehat{\mathbf{Z}}_{[t-1]} \right)
\end{aligned}$$

with

$$\begin{aligned}
\hat{\Theta}_t = & \left\{ (\boldsymbol{\xi}_{[t]}, \boldsymbol{\zeta}_{[t]}, \bar{\boldsymbol{\zeta}}_{[t]}, \hat{\boldsymbol{\zeta}}_{[t]}) \in \mathbb{R}^{\tilde{K}_{[t]}} : \sum_{t'=1}^t [\mathbf{W}_{tt'} \boldsymbol{\xi}_{t'} + \overline{\mathbf{W}}_{tt'} \boldsymbol{\zeta}_{t'}] \right. \\
& \left. \leq \sum_{t'=1}^{t-1} \sum_{t''=1}^{t'} \sum_{i \in \widehat{\mathcal{I}}_{t''}} \left[\mathbf{U}_{tt'}^c \left(\overline{\mathbf{X}}_{t't''i}^c \bar{\boldsymbol{\zeta}}_{t''i} + \widehat{\mathbf{X}}_{t't''i}^c \hat{\boldsymbol{\zeta}}_{t''i} \right) + \mathbf{U}_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b \hat{\boldsymbol{\zeta}}_{t''i} + \overline{\mathbf{U}}_{tt'} \widehat{\mathbf{Y}}_{t't''i} \hat{\boldsymbol{\zeta}}_{t''i} + \widehat{\mathbf{U}}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \hat{\boldsymbol{\zeta}}_{t''i} \right] \right\}.
\end{aligned}$$

For ease of exposition, we define new constraint and right-hand-side matrices such that we

can express problem (19) in the following compact form:

$$\begin{aligned}
& \underset{\substack{\bar{\mathbf{X}}^c, \widehat{\mathbf{X}}^c, \\ \widehat{\mathbf{X}}^b, \widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}}}{\text{minimize}} & \mathbf{e}_1^\top (\bar{\mathbf{X}}_{111}^c + \widehat{\mathbf{X}}_{111}^c) \\
& \text{subject to} & \sum_{t'=1}^t \sum_{t''=1}^{t'} \sum_{i \in \widehat{\mathcal{I}}_{t''}} \left[\tilde{\mathbf{A}}_{tt'}^c \bar{\mathbf{X}}_{t't''i}^c \bar{\zeta}_{t''i} + \left(\tilde{\mathbf{A}}_{tt'}^c \widehat{\mathbf{X}}_{t't''i}^c + \tilde{\mathbf{A}}_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b + \tilde{\mathbf{D}}_{tt'} \widehat{\mathbf{Y}}_{t't''i} + \tilde{\mathbf{H}}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \right) \hat{\zeta}_{t''i} \right] \\
& & \leq \sum_{t'=1}^t \tilde{\mathbf{B}}_{tt'} \boldsymbol{\xi}_{t'} \quad \forall t \in \mathcal{T}, (\boldsymbol{\xi}_{[t]}, \boldsymbol{\zeta}_{[t]}, \bar{\boldsymbol{\zeta}}_{[t]}, \hat{\boldsymbol{\zeta}}_{[t]}) \in \tilde{\Theta}_t (\bar{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^b, \widehat{\mathbf{Y}}_{[t-1]}, \widehat{\mathbf{Z}}_{[t-1]}) \\
& & (\bar{\mathbf{X}}_{tt'i}^c, \widehat{\mathbf{X}}_{tt'i}^c, \widehat{\mathbf{X}}_{tt'i}^b, \widehat{\mathbf{Y}}_{tt'i}, \widehat{\mathbf{Z}}_{tt'i}) \in \Omega_{tt'i} \quad \forall t \in \mathcal{T}, t' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t'},
\end{aligned} \tag{20}$$

where $\tilde{\mathbf{A}}_{tt'}^c \in \mathbb{R}^{P_t \times \bar{N}_{t'}}$, $\tilde{\mathbf{A}}_{tt'}^b \in \mathbb{R}^{P_t \times \widehat{N}_{t'}}$, $\tilde{\mathbf{D}}_{tt'} \in \mathbb{R}^{P_t \times |\bar{\mathcal{I}}_{t'}+1|}$, $\tilde{\mathbf{H}}_{tt'} \in \mathbb{R}^{P_t \times |\widehat{\mathcal{I}}_{t'}+1|}$, and $\tilde{\mathbf{B}}_{tt'} \in \mathbb{R}^{P_t \times K_{t'}}$.

6.4 Reformulation

Applying standard robust optimization techniques based on constraint-wise worst-case reformulation and linear programming (LP) duality⁵⁶, the semi-infinite program (20) can be reformulated into the following problem:

$$\begin{aligned}
& \underset{\substack{\bar{\mathbf{X}}^c, \widehat{\mathbf{X}}^c, \widehat{\mathbf{X}}^b, \\ \widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}, \boldsymbol{\Phi}, \boldsymbol{\Psi}}}{\text{minimize}} & \mathbf{e}_1^\top (\bar{\mathbf{X}}_{111}^c + \widehat{\mathbf{X}}_{111}^c) \\
& \text{subject to} & \sum_{t'=1}^t \sum_{i \in \widehat{\mathcal{I}}_{t'}} \psi_{tt'i} \leq \mathbf{0} \quad \forall t \in \mathcal{T} \\
& & \boldsymbol{\Phi}_t \mathbf{w}_{tt'i} = -\tilde{\mathbf{b}}_{tt'i} \quad \forall t \in \mathcal{T}, t' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t'} \\
& & \psi_{tt'i} + \boldsymbol{\Phi}_t \left[\bar{\mathbf{w}}_{tt''i} v_{t''i} - \sum_{t'=t''}^{t-1} \mathbf{U}_{tt'}^c \bar{\mathbf{X}}_{t't''i}^c \bar{\mathbf{v}}_{t''i} \right. \\
& & \quad \left. - \sum_{t'=t''}^{t-1} \left(\mathbf{U}_{tt'}^c \widehat{\mathbf{X}}_{t't''i}^c + \mathbf{U}_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b + \bar{\mathbf{U}}_{tt'} \widehat{\mathbf{Y}}_{t't''i} + \widehat{\mathbf{U}}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \right) \hat{\mathbf{v}}_{t''i} \right] \\
& & \geq \sum_{t'=t''}^t \left[\tilde{\mathbf{A}}_{tt'}^c \bar{\mathbf{X}}_{t't''i}^c \bar{\mathbf{v}}_{t''i} + \left(\tilde{\mathbf{A}}_{tt'}^c \widehat{\mathbf{X}}_{t't''i}^c + \tilde{\mathbf{A}}_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b + \tilde{\mathbf{D}}_{tt'} \widehat{\mathbf{Y}}_{t't''i} + \tilde{\mathbf{H}}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \right) \hat{\mathbf{v}}_{t''i} \right] \\
& & \quad \forall t \in \mathcal{T}, t'' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t''}, \bar{\mathbf{v}}_{t''i} \in \tilde{\mathcal{V}}_{t''i} \\
& & \psi_{tt'i} \in \mathbb{R}^{P_t} \quad \forall t \in \mathcal{T}, t' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t'} \\
& & \boldsymbol{\Phi}_t \in \mathbb{R}_+^{P_t \times Q_t} \quad \forall t \in \mathcal{T} \\
& & (\bar{\mathbf{X}}_{tt'i}^c, \widehat{\mathbf{X}}_{tt'i}^c, \widehat{\mathbf{X}}_{tt'i}^b, \widehat{\mathbf{Y}}_{tt'i}, \widehat{\mathbf{Z}}_{tt'i}) \in \Omega_{tt'i} \quad \forall t \in \mathcal{T}, t' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t'}
\end{aligned} \tag{21}$$

where $\mathbf{w}_{tt'i}$, $\bar{\mathbf{w}}_{tt'i}$, $\tilde{\mathbf{b}}_{tt'i}$, and ψ_{ti} denote the columns of $\mathbf{W}_{tt'}$, $\bar{\mathbf{W}}_{tt'}$, $\tilde{\mathbf{B}}_{tt'}$, and $\boldsymbol{\Psi}_t$, respectively, corresponding to uncertain parameter i . The detailed derivation of the reformulation can be found in the supplementary material.

Problem (21) is generally a mixed-integer quadratically constrained program (MIQCP) as it involves bilinear terms. However, if the endogenous uncertainty set only depends on binary vari-

ables, the problem can be reformulated into a mixed-integer linear program (MILP) by expressing each integer variable as the difference between two binary variables and linearizing the bilinear terms, which will all be products of a continuous and a binary variable (for more details, see Feng et al. (2020)).

7 Computational Case Studies

In this section, we present the results from four computational case studies that demonstrate the versatility of the proposed modeling framework. All model instances were implemented in Julia v1.3.1 using the modeling environment JuMP v0.21.2⁵⁷ and solved using Gurobi v8.1.1 on a Intel Core i7-8700 CPU at 3.20 GHz machine with 8 GB RAM.

7.1 Pilot Plant and Plant Redesign

We first consider an illustrative example that is based on Example 1. In this problem, we have a design of an industrial-scale plant whose capacity q is uncertain and only becomes known after the plant is built. However, as described in Example 1, a pilot plant can be built such that its capacity p helps provide a more accurate prediction of q . Once p is observed, the marginal uncertainty set for q is updated, based on which we can decide to build one or two plants using the current design or redesign the plant and build a plant using the new design. The solution has to ensure that a given demand d can be met, and the objective is to optimize the worst case. The associated robust optimization problem can be formulated as follows:

$$\begin{aligned}
& \underset{x, y, z_1, z_2}{\text{minimize}} && \delta x + \max_{(p, q, \gamma) \in \Xi(x)} \gamma y(xp) + \rho [z_1(xp) + z_2(xp)] \\
& \text{subject to} && x \in \{0, 1\} \\
& && \left. \begin{aligned} y(xp) &\leq x \\ d &\leq q [z_1(xp) + z_2(xp)] + dy(xp) \\ y(xp), z_1(xp), z_2(xp) &\in \{0, 1\} \end{aligned} \right\} \quad \forall (p, q, \gamma) \in \Xi(x),
\end{aligned} \tag{22}$$

where the binary variable x equals 1 if the pilot plant is built and 0 otherwise, y equals 1 if a plant with a new design and capacity d is built, and z_1 and z_2 indicate how many (maximum two) plants with the current design are built. Nonnegative cost coefficients in the cost function are denoted by δ , γ , and ρ , where γ is an uncertain parameter as it depends on the pilot plant capacity. As indicated by the constraint $y(xp) \leq x$, a plant redesign is only possible if a pilot plant is built and hence more process knowledge is available. Finally, the total plant capacity has to be greater than or equal to demand d . The uncertainty set depends on x and can be expressed as follows:

$$\Xi(x) = \left\{ (p, q, \gamma) \in \mathbb{R}_+^3 : \begin{aligned} & p^{\min}x \leq p \leq p^{\max}x \\ & q^{\min}(1-x) + (\alpha - \Delta)x + \beta p \leq q \leq q^{\max}(1-x) + (\alpha + \Delta)x + \beta p \\ & \gamma = (\rho + \tilde{\gamma}p^{\max})x - \tilde{\gamma}p \end{aligned} \right\},$$

where p and γ can only take nonzero values if $x = 1$. Similarly, $(\alpha - \Delta) + \beta p \leq q \leq (\alpha + \Delta) + \beta p$ if $x = 1$, and $q^{\min} \leq q \leq q^{\max}$ if $x = 0$. The parameters are given such that $q^{\min} = (\alpha - \Delta) + \beta p^{\min}$ and $q^{\max} = (\alpha + \Delta) + \beta p^{\max}$. We further assume that $q^{\min} < d \leq 2q^{\min}$ and that there exists a $\hat{p} < p^{\max}$ such that $d = (\alpha - \Delta) + \beta \hat{p}$.

7.1.1 Analytical Solution

Problem (22) can be solved analytically, and the solution is illustrated in the decision diagram shown in Figure 4. If we choose to build a pilot plant ($x = 1$), we can take recourse actions (y , z_1 , and z_2) based on the realization of p . If $p \geq \hat{p}$, we know that $q \geq d$; hence, we only have to build one plant using the current design. If $p < \hat{p}$, building one plant with the current design may not be sufficient to meet demand d . In this case, we choose the less expensive option between building one plant with a new design that achieves a capacity of d and building two plants with the current design. The costs for these two choices are $v = \delta + \rho + \tilde{\gamma}(p^{\max} - p)$ and $v = \delta + 2\rho$, respectively. Thus, the first option is less expensive than the second if $\tilde{\gamma}(p^{\max} - p) < \rho$. In contrast, if we do not build a pilot plant ($x = 0$), there will be no recourse; in that case, in order to guarantee feasibility for every possible realization of q , we have to build two plants using the current design, which results in a cost of $v = 2\rho$.

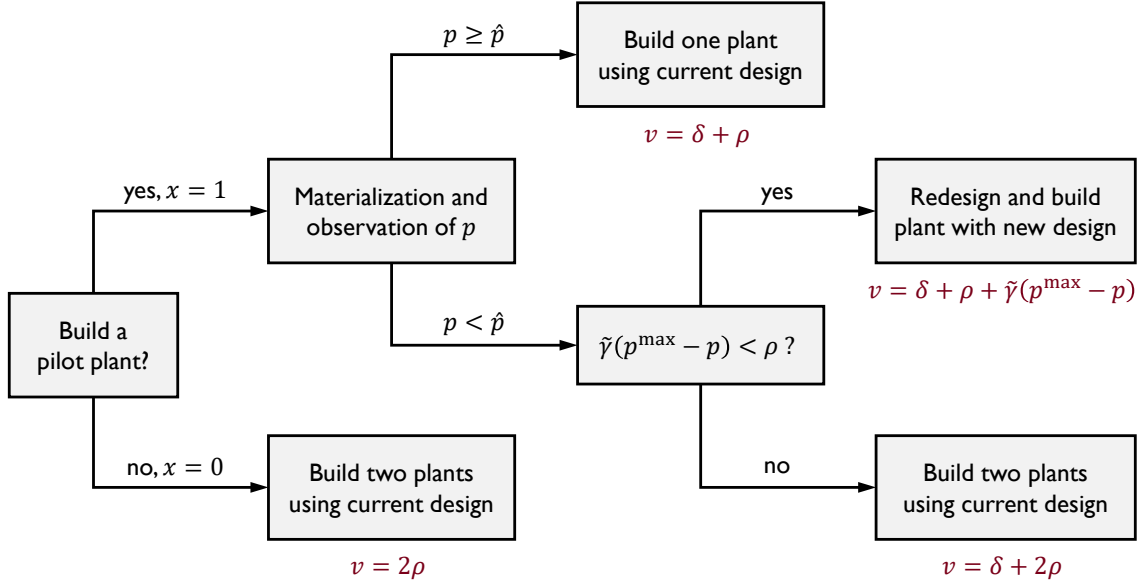


Figure 4: Decision diagram for illustrative pilot plant example.

The problem further simplifies since the objective is to minimize the worst-case cost and the worst case is already known a priori, namely $p = p^{\min}$, $q = q^{\min}$, and $\gamma = \rho + \tilde{\gamma}(p^{\max} - p^{\min})$. Therefore, the optimal value of (22) is simply

$$v^* = \min \{2\rho, \delta + \rho + \tilde{\gamma}(p^{\max} - p^{\min})\},$$

and the corresponding optimal solution is as follows: If $\rho < \delta + \tilde{\gamma}(p^{\max} - p^{\min})$, $x^* = y^* = 0$ and $z_1^* = z_2^* = 1$; if $\rho > \delta + \tilde{\gamma}(p^{\max} - p^{\min})$, $x^* = y^* = 1$ and $z_1^* = z_2^* = 0$. Both solutions are optimal if

$$\rho = \delta + \tilde{\gamma}(p^{\max} - p^{\min}).$$

For our computational case study, we choose $\rho = 100$, $d = 59$, $p^{\min} = 5$, $p^{\max} = 10$, $q^{\min} = 35$, $q^{\max} = 75$, $\alpha = 10$, $\beta = 6$, and $\Delta = 5$. The heat map in Figure 5a shows v^* as a function of δ and $\tilde{\gamma}$. The red line is given by $\delta = \rho - (p^{\max} - p^{\min})\tilde{\gamma} = 100 - 5\tilde{\gamma}$, which indicates the boundary at which v^* switches from $\delta + \rho + \tilde{\gamma}(p^{\max} - p^{\min})$ to 2ρ .

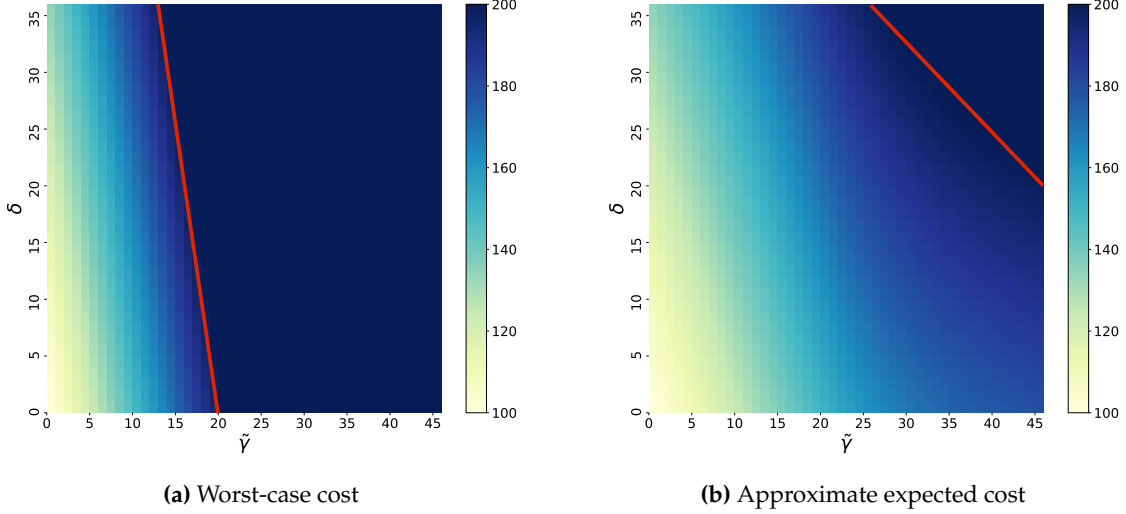


Figure 5: The heat maps show the (a) worst-case cost and (b) scenario-based approximation of the expected cost as functions of δ and $\tilde{\gamma}$.

7.1.2 Non-Worst-Case Objective Function

As mentioned above, in this example, recourse is only possible if $x = 1$, which essentially means that this decision determines whether the problem is static or two-stage. However, note that recourse is not required in the optimal solution of (22), which is due to the fact that worst-case cost minimization is considered and that the worst case is known a priori. In the following, we consider an alternative, scenario-based objective function, which is often a better choice in practice. Here, we introduce a discrete scenario set \mathcal{S} , which contains possible realizations of p , each denoted by \bar{p}_s with probability ϕ_s , where $\sum_{s \in \mathcal{S}} \phi_s = 1$. We then replace the objective function in problem (22) with

$$\delta x + \sum_{s \in \mathcal{S}} \phi_s (\bar{\gamma}_s y(x\bar{p}_s) + \rho [z_1(x\bar{p}_s) + z_2(x\bar{p}_s)]) \quad (23)$$

with $\bar{\gamma}_s = \rho + \tilde{\gamma}(p^{\max} - \bar{p}_s)$. This objective function resembles a sample average approximation of the expected cost as commonly used in stochastic programming. We further reformulate the constraints to eliminate random recourse and remove γ from the uncertainty set, arriving at the

following formulation:

$$\begin{aligned}
& \underset{x, y, z_1, z_2, w_1, w_2}{\text{minimize}} && \delta x + \sum_{s \in \mathcal{S}} \phi_s (\bar{\gamma}_s y(x \bar{p}_s) + \rho [z_1(x \bar{p}_s) + z_2(x \bar{p}_s)]) \\
& \text{subject to} && x \in \{0, 1\} \\
& && \left. \begin{aligned}
& y(xp) \leq x \\
& w_1(xp) \leq q \\
& w_1(xp) \leq q^{\max} z_1(xp) \\
& w_2(xp) \leq q \\
& w_2(xp) \leq q^{\max} z_2(xp) \\
& d \leq w_1(xp) + w_2(xp) + dy(xp) \\
& y(xp), z_1(xp), z_2(xp) \in \{0, 1\} \\
& w_1(xp), w_2(xp) \in \mathbb{R}_+
\end{aligned} \right\} \quad \forall (p, q) \in \bar{\Xi}(x)
\end{aligned} \tag{24}$$

with

$$\bar{\Xi}(x) = \left\{ (p, q) \in \mathbb{R}_+^2 : \begin{aligned} & p^{\min} x \leq p \leq p^{\max} x \\ & q^{\min}(1-x) + (\alpha - \Delta)x + \beta p \leq q \leq q^{\max}(1-x) + (\alpha + \Delta)x + \beta p \end{aligned} \right\}.$$

Note that we have introduced two new continuous recourse variables, w_1 and w_2 , which represent the production amounts of plants built with the current design.

The optimal solution to this modified problem is less trivial as now recourse decisions, as depicted in Figure 4, play a crucial role. Still, we can solve it analytically. If we choose to build a pilot plant, the optimal recourse decisions are just as described in Section 7.1.1. As a result, the optimal value is

$$v^* = \min \left\{ 2\rho, \delta + \sum_{s \in \mathcal{S}_1} \phi_s \rho + \sum_{s \in \mathcal{S}_2} \phi_s [\rho + \tilde{\gamma}(p^{\max} - \bar{p}_s)] + \sum_{s \in \mathcal{S}_3} 2\phi_s \rho \right\},$$

where $\mathcal{S}_1 := \{s \in \mathcal{S} : \bar{p}_s \geq \hat{p}\}$, $\mathcal{S}_2 := \{s \in \mathcal{S} : \bar{p}_s < \hat{p}, \tilde{\gamma}(p^{\max} - \bar{p}_s) < \rho\}$, and $\mathcal{S}_3 := \{s \in \mathcal{S} : \bar{p}_s < \hat{p}, \tilde{\gamma}(p^{\max} - \bar{p}_s) \geq \rho\}$ such that $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$.

We can also apply the proposed decision rule approach to solve this problem. To obtain the exact optimal solution, the marginal support of p needs to be partitioned using appropriate breakpoints. Depending on the parameter values, we require at least one or two breakpoints. If $\tilde{\gamma}(p^{\max} - p^{\min}) \leq \rho$ or $\tilde{\gamma}(p^{\max} - \hat{p}) \geq \rho$, we only need one breakpoint at \hat{p} . If $\tilde{\gamma}(p^{\max} - p^{\min}) > \rho$ and $\tilde{\gamma}(p^{\max} - \hat{p}) < \rho$, we need two breakpoints, one at $p^{\max} - \rho/\tilde{\gamma}$ and one at \hat{p} . Note that the same decision rules are used across all scenarios in \mathcal{S} ; as a result, compared to the worst-case formulation, no additional variables are added to the model such that a comparable computational complexity is achieved. We solve the problem using 100 scenarios sampled from a uniform distribution supported on the given uncertainty set, and obtain the heat map shown in Figure 5b.

7.2 Maintenance Planning with Inspections

This example involves type-1, type-2a, and type-2b endogenous uncertainty. Here, we consider integrated production and maintenance planning with equipment degradation. We use the remaining useful life (RUL) as a measure of equipment health, which decreases with continued operation. To avoid equipment failure, the RUL has to remain greater than zero and can be restored through maintenance. Investing in an equipment upgrade, which would reduce the degradation rate, is also considered. However, degradation processes are inherently stochastic, which renders the change in RUL an uncertain parameter. Furthermore, RUL is an abstract quantity that is often difficult to estimate from operational process data such that elaborate inspections are required to obtain reliable equipment health information. Therefore, in this example, we assume that the RUL can only be observed through inspections, and formulate the resulting multistage robust optimization problem as follows:

$$\begin{aligned}
& \underset{x, \mathbf{y}, \mathbf{p}, \mathbf{s}, \mathbf{r}, \mathbf{m}, \mathbf{f}, \mathbf{z}}{\text{minimize}} && \alpha x + \max_{(\xi, \bar{\xi}, \zeta) \in \Theta_T(x, \mathbf{y}, \mathbf{z})} \sum_{t \in \mathcal{T}} (\beta y_t + \gamma p_t + \delta z_t + \rho m_t - \varphi f_t) \\
& \text{subject to} && x \in \{0, 1\} \\
& && \left. \begin{aligned}
s_t &= s_0 + \sum_{t'=1}^t (p_{t'} - d_{t'}) \leq s^{\max} \\
p_t &\leq C y_t \\
r_t &= r_0 - \sum_{t'=1}^t \xi_{t'} + \sum_{t'=1}^t f_{t'} \leq r^{\max} \\
f_t &\leq r^{\max} m_t \\
y_t + m_t &\leq 1 \\
m_t &\leq z_{t-1} \\
p_t, s_t, r_t, f_t &\in \mathbb{R}_+, \quad y_t, z_t, m_t \in \{0, 1\}
\end{aligned} \right\} \quad \forall t \in \mathcal{T}, \quad (\xi_{[t]}, \bar{\xi}_{[t]}, \zeta_{[t]}) \in \Theta_t(x, \mathbf{y}_{[t]}, \mathbf{z}_{[t]}), \tag{25}
\end{aligned}$$

where x equals 1 if an equipment upgrade is performed, y_t equals 1 if the plant operates in time period t , p_t is the amount produced in time period t , s_t is the inventory level at time t (end of time period t), r_t is the RUL at time t , m_t equals 1 if maintenance is performed in time period t , f_t is the amount of RUL restored in time period t , and z_t equals 1 if an inspection is conducted at time t . The recourse variables are functions of uncertain parameters; however, for the sake of brevity, we do not explicitly indicate this in (25). With given positive cost coefficients α , β , γ , δ , ρ , and φ , the objective is to minimize the worst-case cost over the planning horizon given by the set of time periods $\mathcal{T} = \{1, \dots, T\}$. Note that a negative cost is assigned to f_t in order to encourage maintenance toward the end of the equipment's residual life and to retain a reasonable RUL value at the end of the planning horizon. Given an initial inventory s_0 and demand d_t for every time period t , the second constraint in (25) expresses material balance and sets an upper bound s^{\max} on the inventory. The next constraint states that the production amount is bounded by the plant capacity C . Assuming the initial RUL r_0 is known, r_t is computed and bounded in the next constraint. The RUL reduction caused by operation in time period t is denoted by ξ_t , which is uncertain. If maintenance is performed, the RUL can be restored to its maximum r^{\max} , as stated in the next inequality. Finally, the last two constraints state that maintenance can only be

performed if the plant is shut down and needs to be immediately preceded by an inspection. We further assume that $z_0 = m_1 = f_1 = 0$.

The decision-dependent uncertainty set, which includes auxiliary uncertain parameters, is

$$\Theta_t(x, \mathbf{y}_{[t]}, \mathbf{z}_{[t]}) = \left\{ (\xi_{[t]}, \bar{\xi}_{[t]}, \zeta_{[t]}) \in \mathbb{R}_+^{3t} : \begin{array}{l} \bar{\xi}_{t'} = \sum_{t''=1}^{t'} \xi_{t''} \quad \forall t' = 1, \dots, t \\ \xi^{\min} y_{t'} \leq \xi_{t'} \leq \xi^{\max} y_{t'} \quad \forall t' = 1, \dots, t \\ \xi_{t'} \leq \xi^{\max} - \xi^\Delta x \quad \forall t' = 1, \dots, t \\ \bar{\xi}_{t'} - t' \xi^{\max} (1 - z_{t'}) \leq \zeta_{t'} \leq \bar{\xi}_{t'} \quad \forall t' = 1, \dots, t \\ \zeta_{t'} \leq \bar{\xi}_{t'}^{\max} z_{t'} \quad \forall t' = 1, \dots, t \end{array} \right\},$$

where the auxiliary uncertain parameter $\bar{\xi}_t$ is the cumulative RUL reduction up to time t . The reason for introducing $\bar{\xi}_t$ is that an inspection at time t reveals $\bar{\xi}_t$ but not each individual $\xi_{t'}$ that has contributed to the cumulative RUL reduction since the last inspection. The RUL reduction in time period t , ξ_t , is zero if the plant does not operate, i.e. $y_t = 0$, and bounded by ξ^{\min} and ξ^{\max} otherwise. This upper bound is further reduced by ξ^Δ if an equipment upgrade is performed, i.e. $x = 1$. We have another auxiliary variable ζ_t for every $t \in \mathcal{T}$, which takes the value of $\bar{\xi}_t$ if it is observed through an inspection at time t , i.e. $z_t = 1$, and is zero otherwise.

Note that \mathcal{T} denotes the set of time periods, not the set of stages. In fact, strictly speaking, the number of stages depends on the solution and is directly related to the number of inspections. The maximum possible number of stages is T , which is only achieved if inspection is performed at every $t \in \mathcal{T}$. In that case, after eliminating the state variables s_t and r_t , the first-stage decisions are x, y_1, p_1 , and z_1 . Then, $\bar{\xi}_1$ is observed and the second-stage decisions are y_2, p_2, m_2, f_2 , and z_2 ; this continues similarly up to stage T . The number of stages is implicitly encoded in the recourse variables that are functions of only observed uncertain parameters. Specifically, for all $t \in \mathcal{T}$, $y_t = y_t(\zeta_{[t-1]})$, $p_t = p_t(\zeta_{[t-1]})$, $m_t = m_t(\zeta_{[t-1]})$, $f_t = f_t(\zeta_{[t-1]})$, and $z_t = z_t(\zeta_{[t-1]})$.

We conduct a case study with the following data: $T = 7$, $\beta = \$2 \times 10^3$, $\gamma = \$400$, $\delta = \$10^3$, $\rho = \$2.5 \times 10^5$, $\varphi = \$2 \times 10^3$, $s_0 = 10$, $s^{\max} = 50$, $C = 90$, $r_0 = 40$, $r^{\max} = 125$, $d_t = 50$ for all $t \in \mathcal{T}$, $\xi^{\min} = 4$, $\xi^{\max} = 25$, and $\xi^\Delta = 13$. Two cases are considered: Case 1, in which the equipment upgrade cost α is $\$5 \times 10^4$, and Case 2 with α being $\$10^4$. For each case, we solve one instance using the above ARO model with one breakpoint placed at 8 for each ξ_t , and another instance that disallows recourse, resulting in a static problem (SRO). The optimal values and solution times for all four instances are shown in Table 2. In both Cases 1 and 2, one can observe a clear benefit from recourse, which is reflected in an improvement in the optimal value of about 3%. However, the computational cost for solving the ARO model is significantly higher due to the larger model size. Specifically, the ARO model has 218 binary variables, 320,150 continuous variables, and 952,374 constraints after Gurobi's preprocessing.

The results in Table 2 also indicate that an equipment upgrade can be beneficial if the cost is sufficiently low as the solution suggests investing in an equipment upgrade in Case 2 but not in Case 1. Figure 6 further shows how the equipment upgrade affects the optimal schedule. As depicted in Figure 6a, without equipment upgrade, inspection is performed at time 1. Then, depending on the realization of $\bar{\xi}_1$, the top or bottom schedule is chosen. If $\bar{\xi}_1 \geq 8$, maintenance is performed in time period 2, immediately after the inspection; otherwise, maintenance is post-

Table 2: Computational results for instances of the maintenance planning problem.

	Cost of upgrade (k\$)	Model	Optimal value (k\$)	Solution time (s)
Case 1	50	SRO	219	425
		ARO	212	13,621
Case 2	10	SRO	213	382
		ARO	206	7,837

poned to time period 3. In contrast, the solution for Case 2, as shown in Figure 6b, suggests inspecting the equipment at time 2. This later time point is made possible by the equipment upgrade that reduces the maximum degradation rate. Then, if $\bar{\xi}_2 \geq 16$, maintenance is performed in time period 3; otherwise, maintenance is scheduled for time period 6.

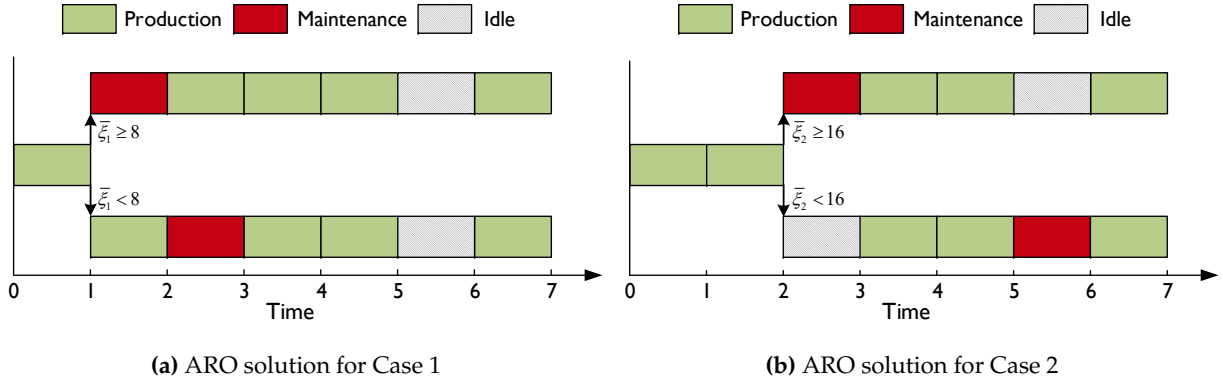


Figure 6: The Gantt charts show the schedules obtained from solving the ARO model for (a) Case 1 with high and (b) Case 2 with low cost of equipment upgrade.

7.3 Optimizing Revision Points in Capacity Planning

The previous example considers a problem in which the number of stages is affected by decisions due to the decision-dependent observation of uncertain parameters. There are other situations in which uncertain parameters are observed over time, independent of decisions, yet we still would like to restrict or optimize the number of stages. One example is long-term capacity planning, which involves large investments that require commitments and lead time for establishing contracts and infrastructure. Hence, it is often infeasible to change decisions frequently despite new information being constantly revealed over time. Instead, one has to restrict oneself to a few revision points at which the planning decisions are revised and updated. In their recent work, Basciftci et al.⁵⁸ have formally introduced an *adaptive two-stage stochastic programming* approach in which for each decision policy, one revision point is considered and chosen as part of the optimization problem. This problem can be viewed as a stochastic program with type-2b endogenous uncertainty where planning decisions for time periods before the revision point have to be made

here and now while decisions after the revision point can be adjusted based on the uncertainty revealed up to the revision point. As a result, our framework can be directly applied to solve a robust optimization variant of the problem and extensions that consider multiple revision points.

Consider a simplified power generation capacity expansion problem with T time periods and a set of types of generation resources \mathcal{I} . At the beginning of each time period t (i.e. at time $t - 1$), the uncertain electricity demand for the same time period, d_t , is revealed and the amount of electricity generated from each generation type i in time period t , y_{it} , is determined. Capacity expansion decisions, however, can only be made or updated at revision points. Let x_{it} denote the amount of generation capacity of type i that is added in time period t and becomes available at time t . Assume that for each generation type i , there are two revision points, one at time $t'_i - 1$, immediately after $d_{t'_i}$ is revealed, and the other one at time $t''_i - 1$ with $t'_i < t''_i$. Then, all x_{it} with $t < t'_i$ have to be chosen before any demand for the planning horizon is known, all x_{it} with $t'_i \leq t < t''_i$ are chosen at time $t'_i - 1$ with known $\mathbf{d}_{[t'_i]}$, and all x_{it} with $t \geq t''_i$ are chosen at time $t''_i - 1$ with known $\mathbf{d}_{[t''_i]}$. Given a maximum number of revision points for each generation type i , denoted by n_i , we aim to place the revision points such that a given cost function, in this case a sample average approximation of the expected cost, is minimized.

We introduce binary variables \hat{z}_{it} , which equals 1 if capacity expansion decisions for generation type i are revised at time $t - 1$, and $z_{itt'}$, which equals 1 if $d_{t'}$ is known when the decision x_{it} is made. The problem can then be formulated as follows:

$$\underset{\mathbf{z}, \hat{\mathbf{z}}, \mathbf{x}, \mathbf{y}}{\text{minimize}} \quad \sum_{i \in \mathcal{I}} \alpha_i x_{i0} + \sum_{s \in \mathcal{S}} \phi_s \sum_{t \in \mathcal{T}} \frac{1}{(1+r)^t} \left[\sum_{i \in \mathcal{I}} (\alpha_i \bar{x}_{its} + \beta_i \bar{y}_{its}) \right] \quad (26a)$$

$$\text{subject to} \quad \sum_{t \in \mathcal{T}} \hat{z}_{it} \leq n_i \quad \forall i \in \mathcal{I} \quad (26b)$$

$$\frac{1}{n_i} \sum_{t''=t'}^t \hat{z}_{it''} \leq z_{itt'} \leq \sum_{t''=t'}^t \hat{z}_{it''} \quad \forall i \in \mathcal{I}, t \in \mathcal{T}, t' = 1, \dots, t \quad (26c)$$

$$\hat{z}_{it} \in \{0, 1\} \quad \forall i \in \mathcal{I}, t \in \mathcal{T} \quad (26d)$$

$$z_{itt'} \in \{0, 1\} \quad \forall i \in \mathcal{I}, t \in \mathcal{T}, t' = 1, \dots, t \quad (26e)$$

$$x_{it} \leq C_{it} \quad \forall i \in \mathcal{I}, t \in \mathcal{T} \quad (26f)$$

$$y_{it} \leq \sum_{t'=1}^t x_{i,t'-1} \quad \forall i \in \mathcal{I}, t \in \mathcal{T}, (\mathbf{d}_{[t]}, \boldsymbol{\zeta}_{[t]}) \in \Theta_t(\mathbf{z}_{[t]}) \quad (26g)$$

$$\sum_{i \in \mathcal{I}} y_{it} \geq d_t \quad \forall t \in \mathcal{T}, (\mathbf{d}_{[t]}, \boldsymbol{\zeta}_{[t]}) \in \Theta_t(\mathbf{z}_{[t]}) \quad (26h)$$

$$x_{i0} \in \mathbb{R}_+ \quad \forall i \in \mathcal{I} \quad (26i)$$

$$y_{it}, x_{it} \in \mathbb{R}_+ \quad \forall i \in \mathcal{I}, t \in \mathcal{T}, (\mathbf{d}_{[t]}, \boldsymbol{\zeta}_{[t]}) \in \Theta_t(\mathbf{z}_{[t]}), \quad (26j)$$

where $\mathcal{T} = \{1, \dots, T\}$, \mathcal{S} denotes the set of scenarios, α_i and β_i are cost coefficients, and r is a discount factor. Each scenario $s \in \mathcal{S}$ is characterized by a specific electricity demand profile $\bar{\mathbf{d}}_{[T],s}$ and probability ϕ_s . In the objective function (26a), $\bar{x}_{its} = x_{it}(\mathbf{z}_{it} \circ \bar{\mathbf{d}}_{[t],s})$ and $\bar{y}_{its} = y_{it}(\bar{\mathbf{d}}_{[t],s})$ represent the recourse decisions for specific scenarios. Constraints (26b) restrict the number of revision points. Inequalities (26c) force $z_{itt'}$ to be 1 if there is at least one revision point between

times $t' - 1$ and $t - 1$ and 0 otherwise. The rationale is that if there is one or multiple revision points between $t' - 1$ and $t - 1$, then x_{it} will be determined at the latest revision point within that time frame at which point $d_{t'}$ will be known. Constraints (26f) set an upper bound, C_{it} , on each x_{it} . Capacity constraints are given in (26g), and constraints (26h) state that the total amount of electricity produced has to meet or exceed the demand. The decision-dependent uncertainty set is

$$\Theta_t(z_{[t]}) = \left\{ (d_{[t]}, \zeta_{[t]}) \in \mathbb{R}_+^{K_t} : \begin{array}{l} d_1^{\min} \leq d_1 \leq d_1^{\max} \\ \delta_{t'}^{\min} d_{t'-1} \leq d_{t'} \leq \delta_{t'}^{\max} d_{t'-1} \quad \forall t' = 2, \dots, t \\ \sum_{t''=1}^{t'} d_{t''} \leq \sum_{t''=1}^{t'} \bar{\delta}^{t''-1} d_1 \quad \forall t' = 1, \dots, t \\ d_{t''} - d_{t''}^{\max} (1 - z_{it't''}) \leq \zeta_{it't''} \leq d_{t''} \quad \forall i \in \mathcal{I}, t' = 1, \dots, t, t'' = 1, \dots, t' \\ d_{t''}^{\min} z_{it't''} \leq \zeta_{it't''} \leq d_{t''}^{\max} z_{it't''} \quad \forall i \in \mathcal{I}, t' = 1, \dots, t, t'' = 1, \dots, t' \end{array} \right\},$$

which encodes the correlation between the electricity demands in different time periods as the range in which $d_{t'}$ can vary depends on $d_{t'-1}$ and the budget of uncertainty imposed on $\sum_{t''=1}^{t'} d_{t''}$ is $\sum_{t''=1}^{t'} \bar{\delta}^{t''-1} d_1$, which is a function of d_1 with a given parameter $\bar{\delta}$. The auxiliary uncertain parameters $\zeta_{[t]}$ depend on the choice of revision points and hence on $z_{[t]}$. Note that the vector $z_{[t]}$ contains all $z_{it't''}$ with $i \in \mathcal{I}$, $t' = 1, \dots, t$, and $t'' = 1, \dots, t'$. Also, $K_t = t + |\mathcal{I}|t(t+1)/2$. The recourse variables are adjustable given the following dependence on uncertain parameters: $x_{it} = x_{it}(\zeta_{[t]})$ and $y_{it} = y_{it}(d_{[t]})$.

Notice that z and \hat{z} , which determine the revision points, are all first-stage variables. As a result, the decision-dependent nonanticipativity for the capacity expansion decisions x can alternatively be modeled by forcing the decision rule coefficients associated with unobserved uncertain parameters to zero (see Section 6.1). The decision rules with respect to the lifted uncertain parameters are

$$x_{it} = \sum_{t'=0}^t \left(\bar{x}_{itt'}^\top \bar{d}_{t'} + \hat{x}_{itt'}^\top \hat{d}_{t'} \right) \quad \forall i \in \mathcal{I}, t \in \mathcal{T}$$

with $\bar{d}_0 = 1$ and $\hat{d}_0 = 0$. Decision-dependent nonanticipativity is then enforced by the following constraints:

$$\begin{aligned} -M z_{itt'} e &\leq \bar{x}_{itt'} \leq M z_{itt'} e \quad \forall i \in \mathcal{I}, t \in \mathcal{T}, t' = 1, \dots, t \\ -M z_{itt'} e &\leq \hat{x}_{itt'} \leq M z_{itt'} e \quad \forall i \in \mathcal{I}, t \in \mathcal{T}, t' = 1, \dots, t \end{aligned}$$

with M being a sufficiently large parameter. This alternative approach allows us to work with a fixed uncertainty set:

$$\Theta_t = \left\{ d_{[t]} \in \mathbb{R}_+^t : \begin{array}{l} d_1^{\min} \leq d_1 \leq d_1^{\max} \\ \delta_{t'}^{\min} d_{t'-1} \leq d_{t'} \leq \delta_{t'}^{\max} d_{t'-1} \quad \forall t' = 2, \dots, t \\ \sum_{t''=1}^{t'} d_{t''} \leq \sum_{t''=1}^{t'} \bar{\delta}^{t''-1} d_1 \quad \forall t' = 1, \dots, t \end{array} \right\}.$$

Moreover, it does not require auxiliary uncertain parameters, which significantly reduces the computational complexity.

In the following case study, we consider a planning horizon of seven time periods and four generation types, namely coal, natural gas-gas turbine (NG-GT), solar PV, and wind. Cost data are adapted from Min et al.⁵⁹, which, along with all other required data, can be found in the

supplementary material. Furthermore, the objective function is constructed using 100 scenarios sampled from a uniform distribution supported on the given uncertainty set, and uncertainty lifting is performed with two equidistantly placed breakpoints for each uncertain parameter.

We solve the capacity planning problem for different maximum allowed number of revision points \bar{n} , which is applied to all generation types, i.e. $n_i = \bar{n}$ for all $i \in \mathcal{I}$. The results are shown in Table 3. As expected, the optimal value decreases with increasing number of revision points. The improvement is most significant from the static case (no revision) to the case with one revision point. Diminishing returns are observed at higher number of revision points. The results show that the locations of the revision points for different generation types are generally not the same. For larger \bar{n} , there also seems to be no need to use the maximum number of revision points for all generation types. Finally, note that the solution times are moderate, indicating the potential to solve more complex problems of this kind using the proposed approach.

Table 3: Computational results for the capacity planning problem with different allowed numbers of revision points. The relative improvement is computed with respect to the optimal value of the case with no revision.

Allowed # of revision points	Optimal value (k\$)	Solution time (s)	Relative improvement	Location of revision points			
				Coal	NG-GT	Solar PV	Wind
0	46,183	9					
1	40,852	427	11.54 %	3	5	1	2
2	39,526	1,679	14.42 %	3, 6	2, 5	1, 6	1, 3
3	38,955	764	15.65 %	2, 3, 6	1, 2, 5	1, 5, 6	1, 3, 5
4	38,612	874	16.39 %	2, 3, 5, 6	1, 2, 3, 5	1, 5, 6	1, 3, 5
5	38,423	592	16.80 %	1, 2, 3, 5, 6	1, 2, 3, 4, 5	5, 6	1, 2, 3, 5
6	38,423	462	16.80 %	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5	5, 6	1, 2, 3, 5

7.4 Production Scheduling with Active Parameter Estimation

When operating a manufacturing process, we often do not have accurate information across the whole range of feasible operation due to the lack of process data. A model that we use for process optimization may only be accurate around one or a few nominal points, especially if it is a relatively new process. For example, a production unit may be able to operate in different operating modes, but process parameters such as efficiency and processing time are only accurately known for the one operating mode that is used most of the time. The models of other, potentially better-performing operating modes can only be improved by actually operating in those modes and collecting more data; however, this exploration action comes with the risk of operating under conditions with inferior performance or damaging impact. In the following example, we consider process scheduling with active model parameter estimation, which integrates scheduling optimization and design of experiments in order to find an optimal trade-off between exploration and exploitation that maximizes the overall plant performance over time. The problem is formulated

as follows:

$$\begin{aligned}
& \underset{\mathbf{x}, \mathbf{s}}{\text{minimize}} && \max_{\mathbf{p} \in \mathcal{P}_T(\mathbf{x})} \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_i} \mathbf{c}^\top \mathbf{u}_{ij} x_{tij}(\mathbf{p}_{[t]}) + \delta s_t(\mathbf{p}_{[t]}) \\
& \text{subject to} && s_t(\mathbf{p}_{[t]}) = s_0 + \sum_{t'=1}^t \left(\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_i} p_{t'ij} - d_{t'} \right) \leq s^{\max} \quad \forall t \in \mathcal{T}, \mathbf{p}_{[t]} \in \mathcal{P}_t(\mathbf{x}_{[t]}) \\
& && \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_i} x_{tij}(\mathbf{p}_{[t]}) \leq 1 \quad \forall t \in \mathcal{T}, \mathbf{p}_{[t]} \in \mathcal{P}_t(\mathbf{x}_{[t]}) \\
& && s_t(\mathbf{p}_{[t]}) \in \mathbb{R}_+ \quad \forall t \in \mathcal{T}, \mathbf{p}_{[t]} \in \mathcal{P}_t(\mathbf{x}_{[t]}) \\
& && x_{tij}(\mathbf{p}_{[t]}) \in \{0, 1\} \quad \forall t \in \mathcal{T}, i \in \mathcal{I}, j \in \mathcal{J}_i, \mathbf{p}_{[t]} \in \mathcal{P}_t(\mathbf{x}_{[t]}),
\end{aligned} \tag{27}$$

where we assume that the plant operating point is chosen among a discrete set of operating points, \mathcal{J}_i , which depends on operating mode i . Each operating point j of mode i is defined by a vector of inputs, \mathbf{u}_{ij} . The binary variable x_{tij} equals 1 if the plant operates at point j of mode i in time period t . The amount produced, p_{tij} , is treated as an uncertain parameter since the model parameters for the corresponding mode i may be uncertain. Given cost coefficients \mathbf{c} and δ , the objective is to minimize the worst-case cost over the entire scheduling horizon $\mathcal{T} = \{1, \dots, T\}$. With an initial inventory s_0 , the inventory level at time t , s_t , is bounded between zero and s^{\max} . The product demand in time period t is denoted by d_t .

We assume that the production amount in each operating mode i is given by a linear function $p = a_i + \mathbf{b}_i^\top \mathbf{u}$. For some modes, the model parameters a_i and \mathbf{b}_i are unknown at the beginning of the scheduling horizon, and they can only be estimated by operating in those modes and observing the resulting production amounts. This is captured in the following decision-dependent uncertainty set:

$$\mathcal{P}_t(\mathbf{x}_{[t]}) = \left\{ \begin{array}{l} \exists a_i, \mathbf{b}_i \forall i \in \mathcal{I} \quad \text{such that} \\ 0 \leq p_{t'ij} \leq p_i^{\max} x_{t'ij} \quad \forall t' = 1, \dots, t, i \in \mathcal{I}, j \in \mathcal{J}_i \\ \mathbf{p}_{[t]} \in \mathbb{R}_+^{K_t} : a_i + \mathbf{b}_i^\top \mathbf{u}_{ij} - p_i^{\max}(1 - x_{t'ij}) \leq p_{t'ij} \leq a_i + \mathbf{b}_i^\top \mathbf{u}_{ij} \quad \forall t' = 1, \dots, t, i \in \mathcal{I}, j \in \mathcal{J}_i \\ a_i^{\min} \leq a_i \leq a_i^{\max} \quad \forall i \in \mathcal{I} \\ \mathbf{b}_i^{\min} \leq \mathbf{b}_i \leq \mathbf{b}_i^{\max} \quad \forall i \in \mathcal{I} \end{array} \right\}$$

with $K_t = t \sum_{i \in \mathcal{I}} |\mathcal{J}_i|$. Only if $x_{tij} = 1$, the production amount p_{tij} can only take a nonzero value, which depends on the chosen operating point. Each operating point that has been selected up to time period t activates a linear equation, which can be interpreted as a cut that reduces the size of the uncertainty set. As a result, the linear model of an operating mode is learned through the accumulation of observations, which are the production amounts at different operating points. If $\mathbf{u}_{ij} \in \mathbb{R}^n$, and all model parameters of mode i are unknown, one needs to operate in mode i at different operating points at least $n + 1$ times in order to uniquely determine a_i and \mathbf{b}_i .

We use a small instance to illustrate the effect of active learning in this problem. The data are as follows: $T = 4$, $c = 5$, $\delta = 2$, $\mathbf{d} = (32.5, 15, 51, 42.5)$, $s^{\max} = 100$ and $s_0 = 0$. We consider two operating modes and two possible operating points per mode, with each operating point $j \in \mathcal{J}_i$ defined by a scalar input u_{ij} . We have $u_{11} = 10$, $u_{12} = 46$, $u_{21} = 25$, and $u_{22} = 44$. The values of the parameters describing the uncertainty set are shown in Table 4. Here, we assume that the plant has not been operated in any of the two modes prior to the beginning of the scheduling horizon.

Table 4: Parameters for the description of the uncertainty set in the production scheduling with active parameter estimation problem.

	p_i^{\max}	a_i^{\min}	a_i^{\max}	b_i^{\min}	b_i^{\max}
Mode 1	65.2	5	10	1	1.2
Mode 2	56.2	20	21	0.5	0.8

The optimal adjustable production schedule is shown in Figure 7. The solution suggests to first operate at point 1 of mode 2 and choose the subsequent schedule depending on the realization of p_{121} . The schedule at the top is selected if $p_{121} \geq 36.75$, and the bottom schedule is selected if $p_{121} < 36.75$. One can see that in both schedules, the operating points selected in time periods 2 and 3 are the same; the schedules only differ in the operating mode chosen for time period 4. By only observing the production amount at operating point 1 of mode 2, we cannot fully infer the model for mode 2, i.e. determine a_2 and b_2 . However, the observation reduces the uncertainty for mode 2, which allows us to determine in which mode the plant should operate in time period 4.

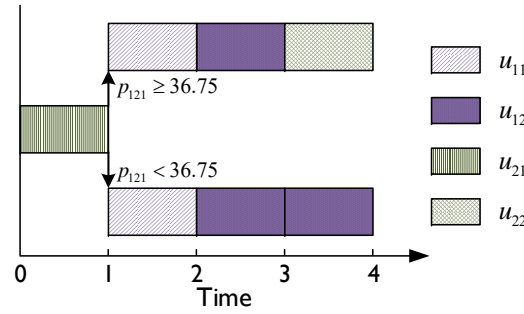


Figure 7: Optimal adjustable production schedule with active parameter estimation. The solution indicates that recourse is only required once.

The effect of learning is further illustrated in Figure 8, which shows how the uncertainty set is reduced over time in a particular scenario with $p_{121} > 36.75$, in which case the schedule at the top of Figure 7 is applied. Each of the four subfigures shows the production amount as a function of the input for a specific time period and the mode chosen for that time period. The black solid vertical lines indicate the marginal uncertainty sets of the production amounts at the available operating points before any observations are made, as given by the ranges of possible values for a_i and b_i . In this particular scenario, a production amount of 39 is observed in time period 1 when the plant operates at point 1 of mode 2. Given this observation and $a_2 \in [20, 21]$, the uncertainty with respect to b_2 is immediately reduced such that the line representing the linear model of mode 2 has to lie within the cyan shaded area shown in Figure 8a. One can see that this results in a substantial reduction of uncertainty with respect to the unobserved operating point 2 of mode 2. The same effect is seen in time period 2 when the plant operates at point 1 of mode 1 (Figure 8b). After operating at point 2 of mode 1 in time period 3, a_1 and b_1 can be uniquely determined given the observations p_{211} and p_{312} . The linear model of mode 1 is represented by the red solid line in Figure 8c. Finally, the linear model of mode 2 will also be exactly known after operating at point 2

of mode 2 in time period 4, as indicated in Figure 8d.

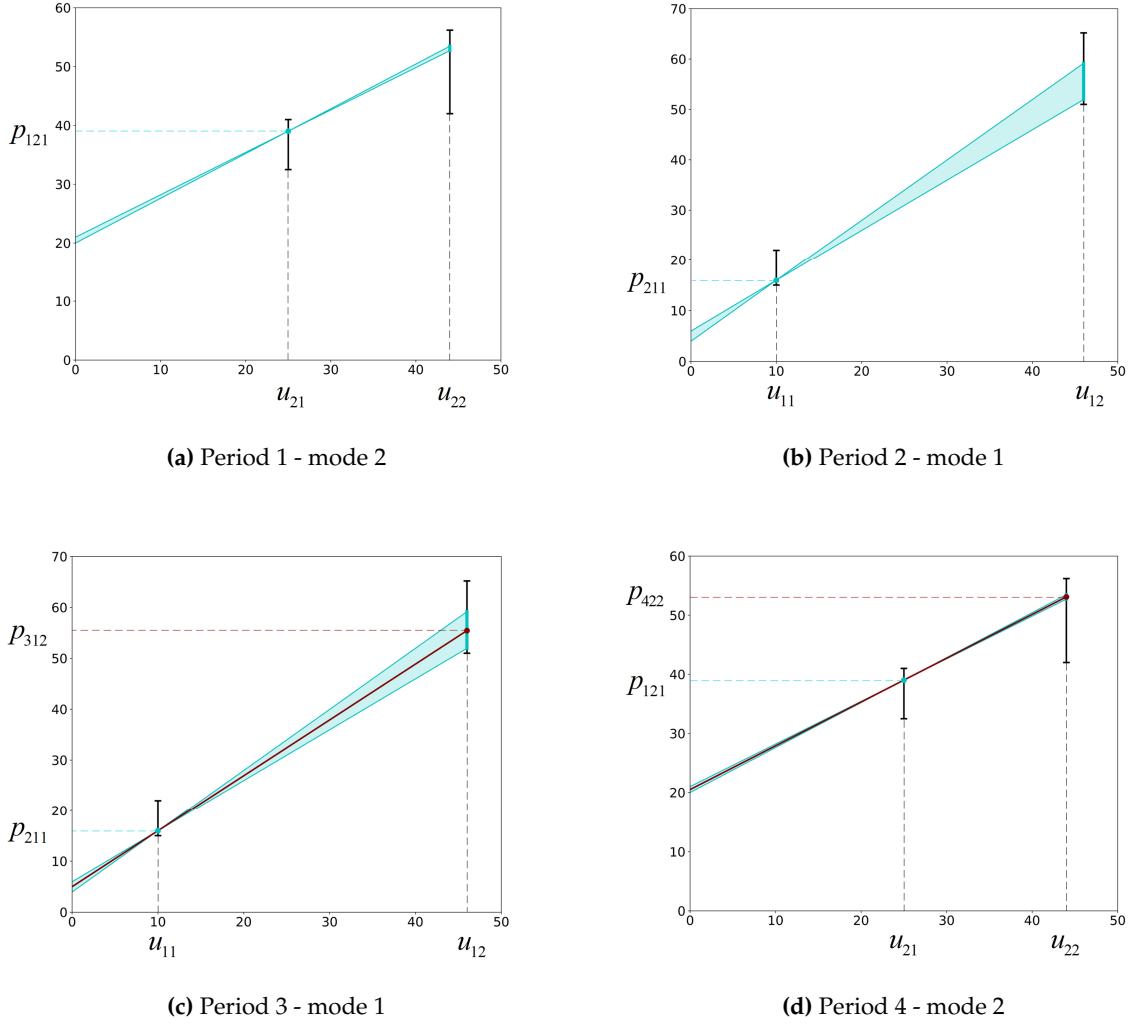


Figure 8: Reduction of uncertainty by observing materialized production amounts over time for one specific scenario (defined by one specific set of observations).

The optimal value of the problem is 819.2. Despite the simplicity of the illustrative example, the problem is computationally challenging. This is in part due to the large number of variables and constraints resulting from the reformulation, but even more due to the weak relaxation of the MILP. The latter is indicated by a lower bound that only improves very slowly in the branch-and-bound algorithm, despite the incumbent being the optimal solution already. Computational strategies for mitigating these challenges will be explored and discussed in future work.

8 Conclusions

In this work, we have developed an adjustable robust optimization framework that can simultaneously consider three types of endogenous uncertainty: type 1, type 2a, and type 2b, as defined

in a new refined classification of endogenous uncertainty that distinguishes between the decision-dependent *materialization* and *observation* of uncertain parameters. One compelling insight from this extended view is that active learning, a sequential form of experimental design, can be formulated as an optimization problem under endogenous uncertainty. As a result, the proposed approach provides a means of performing integrated optimization and active learning in a sequential decision-making environment.

Our framework considers decision-dependent polyhedral uncertainty sets, and we have applied a decision rule approach based on the concept of lifted uncertainty that incorporates continuous and binary recourse, including recourse decisions that affect the uncertainty set. Multistage decision-dependent nonanticipativity is modeled using auxiliary uncertain parameters, which preserves the tractability of the decision rule approach. This eventually results in a mixed-integer optimization problem that can be directly solved using off-the-shelf solvers.

The proposed methodology significantly expands our capability to model and solve data-driven optimization problems under uncertainty, particularly those involving active learning. Computational challenges still exist, which we plan to address in our future work. Nonetheless, this approach promises to be especially suited for applications that involve complex constraints, expensive exploration actions, and strong links between different decision stages. To highlight the relevance of these kinds of problems, we have conducted computational case studies covering a variety of applications. The results demonstrate the versatility of the proposed modeling framework as well as the importance of endogenous uncertainty. As these aspects become increasingly relevant in PSE applications, we hope that this work can contribute toward the development of more comprehensive and efficient methods for such problems.

Acknowledgments

Wei Feng gratefully acknowledges the financial support from the China Scholarship Council (CSC) (No. 201906320317).

References

- [1] A. Ben-Tal and A. Nemirovski. Robust Convex Optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.
- [2] L. El Ghaoui, F. Oustry, and H. Lebret. Robust Solutions to Uncertain Semidefinite Programs. *SIAM Journal on Optimization*, 9(1):33–52, 1998.
- [3] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.
- [4] A. Georghiou, D. Kuhn, and W. Wiesemann. The decision rule approach to optimization under uncertainty: methodology and applications. *Computational Management Science*, 16(4): 545–576, 2019.

- [5] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems*, 28(1):52–63, 2013.
- [6] B. Zeng and L. Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457–461, 2013.
- [7] D. Bertsimas and C. Caramanis. Finite adaptability in multistage linear optimization. *IEEE Transactions on Automatic Control*, 55(12):2751–2766, 2010.
- [8] G. A. Hanasusanto, D. Kuhn, and W. Wiesemann. K-Adaptability in Two-Stage Robust Binary Programming. *Operations Research*, 63(4):877–891, 2015.
- [9] D. Bertsimas and A. Georghiou. Design of Near Optimal Decision Rules in Multistage Adaptive Mixed-Integer Optimization. *Operations Research*, 63(3):610–627, 2015.
- [10] K. Postek and D. den Hertog. Multistage Adjustable Robust Mixed-Integer Optimization via Iterative Splitting of the Uncertainty Set. *INFORMS Journal on Computing*, 28(3):553–574, 2016.
- [11] D. Bertsimas and A. Georghiou. Binary decision rules for multistage adaptive mixed-integer optimization. *Mathematical Programming*, 167(2):395–433, 2018.
- [12] A. Subramanyam, C. E. Gounaris, and W. Wiesemann. K-adaptability in two-stage mixed-integer robust optimization. *Mathematical Programming Computation*, 12:193–224, 2020.
- [13] I. Yanıkoğlu, B. L. Gorissen, and D. den Hertog. A survey of adjustable robust optimization. *European Journal of Operational Research*, 277(3):799–813, 2019.
- [14] N. H. Lappas and C. E. Gounaris. Multi-Stage Adjustable Robust Optimization for Process Scheduling Under Uncertainty. *AIChE Journal*, 62(5):1646–1667, 2016.
- [15] Q. Zhang, M. F. Morari, I. E. Grossmann, A. Sundaramoorthy, and J. M. Pinto. An adjustable robust optimization approach to scheduling of continuous industrial processes providing interruptible load. *Computers and Chemical Engineering*, 86:106–119, 2016.
- [16] H. Shi and F. You. A Computational Framework and Solution Algorithms for Two-Stage Adaptive Robust Scheduling of Batch Manufacturing Processes Under Uncertainty. *AIChE Journal*, 62(3):687–703, 2016.
- [17] Q. Zhang, I. E. Grossmann, and R. M. Lima. On the Relation Between Flexibility Analysis and Robust Optimization for Linear Systems. *AIChE Journal*, 62(9):3109–2132, 2016.
- [18] I. E. Grossmann, B. A. Calfa, and P. Garcia-Herreros. Evolution of concepts and models for quantifying resiliency and flexibility of chemical processes. *Computers and Chemical Engineering*, 70:22–34, 2014.
- [19] Y. Yuan, Z. Li, and B. Huang. Nonlinear robust optimization for process design. *AIChE Journal*, 64(2):481–494, 2018.

- [20] J. Gong and F. You. Resilient design and operations of process systems: Nonlinear adaptive robust optimization model and algorithm for resilience analysis and enhancement. *Computers and Chemical Engineering*, 116:231–252, 2018.
- [21] C. Ning and F. You. A Data-Driven Multistage Adaptive Robust Optimization Framework for Planning and Scheduling Under Uncertainty. *AIChE Journal*, 63(10):4343–4369, 2017.
- [22] N. H. Lappas and C. E. Gounaris. Theoretical and computational comparison of continuous-time process scheduling models for adjustable robust optimization. *AIChE Journal*, 64(8):3055–3070, 2018.
- [23] W. Feng, Y. Zhang, G. Rong, and Y. Feng. Finite Adaptability in Data-Driven Robust Optimization for Production Scheduling: A Case Study of the Ethylene Plant. *Industrial & Engineering Chemistry Research*, 58(16):6505–6518, 2019.
- [24] S. Rahal, Z. Li, and D. J. Papageorgiou. Proactive and reactive scheduling of the steelmaking and continuous casting process through adaptive robust optimization. *Computers and Chemical Engineering*, 133:106658, 2020.
- [25] X. Zhang, M. Kamgarpour, A. Georghiou, P. Goulart, and J. Lygeros. Robust optimal control with adjustable uncertainty sets. *Automatica*, 75:249–259, 2017.
- [26] C. Shang, W. H. Chen, and F. You. Robust constrained model predictive control of irrigation systems based on data-driven uncertainty set constructions. In *Proceedings of the American Control Conference*, pages 1906–1911. American Automatic Control Council, 2019.
- [27] C. Ning and F. You. A transformation-proximal bundle algorithm for multistage adaptive robust optimization and application to constrained robust optimal control. *Automatica*, 113:108802, 2020.
- [28] D. Yue and F. You. Optimal Supply Chain Design and Operations Under Multi-Scale Uncertainties: Nested Stochastic Robust Optimization Modeling Framework and Solution Algorithm. *AIChE Journal*, 62(9):3041–3055, 2016.
- [29] A. Subramanyam, F. Mufalli, J. M. Pinto, and C. E. Gounaris. Robust Multi-Period Vehicle Routing under Customer Order Uncertainty. *Optimization Online*, 2018.
- [30] T. W. Jonsbråten, R. J.-B. Wets, and D. L. Woodruff. A class of stochastic programs with decision dependent random elements. *Annals of Operations Research*, 82:83–106, 1998.
- [31] S. Peeta, F. S. Salman, D. Gunnec, and K. Viswanath. Pre-disaster investment decisions for strengthening a highway network. *Computers and Operations Research*, 37(10):1708–1719, 2010.
- [32] L. F. Escudero, M. A. Garín, J. F. Monge, and A. Unzueta. On preparedness resource allocation planning for natural disaster relief under endogenous uncertainty with time-consistent risk-averse management. *Computers and Operations Research*, 98:84–102, 2018.
- [33] L. Hellemo, P. I. Barton, and A. Tomasgard. Decision-dependent probabilities in stochastic programs with recourse. *Computational Management Science*, 15(3-4):369–395, 2018.

- [34] V. Goel and I. E. Grossmann. A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves. *Computers and Chemical Engineering*, 28(8): 1409–1429, 2004.
- [35] V. Goel and I. E. Grossmann. A class of stochastic programs with decision dependent uncertainty. *Mathematical Programming*, 108:355–397, 2006.
- [36] M. Colvin and C. T. Maravelias. A stochastic programming approach for clinical trial planning in new drug development. *Computers and Chemical Engineering*, 32(11):2626–2642, 2008.
- [37] M. Colvin and C. T. Maravelias. Modeling methods and a branch and cut algorithm for pharmaceutical clinical trial planning using stochastic programming. *European Journal of Operational Research*, 203(1):205–215, 2010.
- [38] P. Vayanos, D. Kuhn, and B. Rustem. Decision rules for information discovery in multi-stage stochastic programming. In *Proceedings of the IEEE Conference on Decision and Control*, pages 7368–7373, 2011.
- [39] B. Tarhan, I. E. Grossmann, and V. Goel. Computational strategies for non-convex multi-stage MINLP models with decision-dependent uncertainty and gradual uncertainty resolution. *Annals of Operations Research*, 203(1):141–166, 2013.
- [40] V. Gupta and I. E. Grossmann. A new decomposition algorithm for multistage stochastic programs with endogenous uncertainties. *Computers and Chemical Engineering*, 62:62–79, 2014.
- [41] B. Christian and S. Cremaschi. Heuristic solution approaches to the pharmaceutical R&D pipeline management problem. *Computers and Chemical Engineering*, 74:34–47, 2015.
- [42] F. Hooshmand and S. A. MirHassani. Efficient constraint reduction in multistage stochastic programming problems with endogenous uncertainty. *Optimization Methods and Software*, 31(2):359–376, 2016.
- [43] R. M. Apap and I. E. Grossmann. Models and computational strategies for multistage stochastic programming under endogenous and exogenous uncertainties. *Computers and Chemical Engineering*, 103:233–274, 2017.
- [44] M. Poss. Robust combinatorial optimization with variable budgeted uncertainty. *4OR*, 11(1): 75–92, 2013.
- [45] O. Nohadani and K. Sharma. Optimization under decision-dependent uncertainty. *SIAM Journal on Optimization*, 28(2):1773–1795, 2018.
- [46] N. H. Lappas and C. E. Gounaris. Robust optimization for decision-making under endogenous uncertainty. *Computers and Chemical Engineering*, 111:252–266, 2018.
- [47] S. Avraamidou and E. N. Pistikopoulos. Adjustable robust optimization through multi-parametric programming. *Optimization Letters*, 14(4):873–887, 2019.
- [48] W. Feng, Y. Feng, and Q. Zhang. Multistage Robust Mixed-Integer Optimization Under Endogenous Uncertainty. *arXiv preprint arXiv:2008.11633*, 2020.

- [49] D. Bertsimas and P. Vayanos. Data-driven learning in dynamic pricing using adaptive optimization. *Optimization Online*, 2014.
- [50] P. Vayanos, A. Georghiou, and H. Yu. Robust optimization with decision-dependent information discovery. *arXiv preprint arXiv:2004.08490*, 2020.
- [51] P. Vayanos, D. C. Mcelfresh, Y. Ye, J. P. Dickerson, and E. Rice. Active Preference Elicitation via Adjustable Robust Optimization. *arXiv preprint arXiv:2003.01899*, 2020.
- [52] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016.
- [53] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [54] J. Y. Audibert, S. Bubeck, and R. Munos. Best arm identification in multi-armed bandits. *COLT 2010 - The 23rd Conference on Learning Theory*, pages 41–53, 2010.
- [55] A. Georghiou, W. Wiesemann, and D. Kuhn. Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming*, 152(1-2):301–338, 2015.
- [56] B. L. Gorissen, I. Yanikoglu, and D. den Hertog. A practical guide to robust optimization. *Omega*, 53:124–137, 2015.
- [57] M. Lubin and I. Dunning. Computing in Operations Research Using Julia. *INFORMS Journal on Computing*, 27(2):237–248, 2015.
- [58] B. Basciftci, S. Ahmed, and N. Gebraeel. Adaptive Two-stage Stochastic Programming with an Application to Capacity Expansion Planning. *arXiv preprint arXiv:1906.03513*, 2019.
- [59] D. Min, J.-h. Ryu, and D. G. Choi. A long-term capacity expansion planning model for an electric power system integrating large-size renewable energy technologies. *Computers and Operations Research*, 96:244–255, 2018.

Supplementary Material to “A Unified Framework for Adjustable Robust Optimization with Endogenous Uncertainty”

Qi Zhang^{*1} and Wei Feng²

¹*Department of Chemical Engineering and Materials Science,
University of Minnesota, Minneapolis, MN 55455, USA*

²*State Key Laboratory of Industrial Control Technology, College of Control Science and
Engineering, Zhejiang University, Hangzhou 310027, China*

1 Detailed Derivation of Reformulation

From the decision-rule-based approximation, we arrive at the following uncertainty set:

$$\begin{aligned} & \tilde{\Theta}_t \left(\bar{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^b, \widehat{\mathbf{Y}}_{[t-1]}, \widehat{\mathbf{Z}}_{[t-1]} \right) \\ &= \left\{ \{\boldsymbol{\xi}_{[t]}\} \times \prod_{t'=1}^t \prod_{i \in \widehat{\mathcal{I}}_{t'}} Z_{t'i} \right\} \cap \hat{\Theta}_t \left(\bar{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^b, \widehat{\mathbf{Y}}_{[t-1]}, \widehat{\mathbf{Z}}_{[t-1]} \right) \end{aligned}$$

with

$$\begin{aligned} \hat{\Theta}_t &= \left\{ (\boldsymbol{\xi}_{[t]}, \boldsymbol{\zeta}_{[t]}, \bar{\boldsymbol{\zeta}}_{[t]}, \hat{\boldsymbol{\zeta}}_{[t]}) \in \mathbb{R}^{\tilde{K}_{[t]}} : \sum_{t'=1}^t [\mathbf{w}_{tt'} \boldsymbol{\xi}_{t'} + \bar{\mathbf{w}}_{tt'} \boldsymbol{\zeta}_{t'}] \right. \\ &\quad \left. \leq \sum_{t'=1}^{t-1} \sum_{t''=1}^{t'} \sum_{i \in \widehat{\mathcal{I}}_{t''}} \left[U_{tt'}^c \left(\bar{\mathbf{X}}_{t't''i}^c \bar{\boldsymbol{\zeta}}_{t''i} + \widehat{\mathbf{X}}_{t't''i}^c \hat{\boldsymbol{\zeta}}_{t''i} \right) + U_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b \hat{\boldsymbol{\zeta}}_{t''i} + \bar{U}_{tt'} \widehat{\mathbf{Y}}_{t't''i} \hat{\boldsymbol{\zeta}}_{t''i} + \hat{U}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \hat{\boldsymbol{\zeta}}_{t''i} \right] \right\} \\ &= \left\{ (\boldsymbol{\xi}_{[t]}, \boldsymbol{\zeta}_{[t]}, \bar{\boldsymbol{\zeta}}_{[t]}, \hat{\boldsymbol{\zeta}}_{[t]}) \in \mathbb{R}^{\tilde{K}_{[t]}} : \sum_{t'=1}^t [\mathbf{w}_{tt'} \boldsymbol{\xi}_{t'} + \bar{\mathbf{w}}_{tt'} \boldsymbol{\zeta}_{t'}] \right. \\ &\quad \left. \leq \sum_{t'=1}^{t-1} \sum_{t''=1}^{t'} \sum_{i \in \widehat{\mathcal{I}}_{t''}} \left[U_{tt'}^c \bar{\mathbf{X}}_{t't''i}^c \bar{\boldsymbol{\zeta}}_{t''i} + \left(U_{tt'}^c \widehat{\mathbf{X}}_{t't''i}^c + U_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b + \bar{U}_{tt'} \widehat{\mathbf{Y}}_{t't''i} + \hat{U}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \right) \hat{\boldsymbol{\zeta}}_{t''i} \right] \right\}. \end{aligned}$$

^{*}Corresponding author (qizh@umn.edu)

Using the definition of $Z_{t'i}$, we can translate $\tilde{\Theta}_t$ into an uncertainty set in the $(\xi_{[t]}, \lambda_{[t]})$ -space, which we denote Λ_t :

$$\begin{aligned}
\Lambda_t = & \left\{ (\xi_{[t]}, \lambda_{[t]}) : \sum_{t'=1}^t \left[\sum_{i \in \tilde{\mathcal{I}}_{t'}} w_{tt'i} \xi_{t'i} + \sum_{i \in \tilde{\mathcal{I}}_{t'}} \bar{w}_{tt'i} \sum_{\tilde{v}_{t'i} \in \tilde{\mathcal{V}}_{t'i}} \lambda_{t'i}(\tilde{v}_{t'i}) v_{t'i} \right] \right. \\
& \leq \sum_{t'=1}^{t-1} \sum_{t''=1}^{t'} \sum_{i \in \hat{\mathcal{I}}_{t''}} \left[U_{tt'}^c \bar{X}_{t't''i}^c \sum_{\tilde{v}_{t''i} \in \tilde{\mathcal{V}}_{t''i}} \lambda_{t''i}(\tilde{v}_{t''i}) \tilde{v}_{t''i} \right. \\
& \quad \left. + \left(U_{tt'}^c \hat{X}_{t't''i}^c + U_{tt'}^b \hat{X}_{t't''i}^b + \bar{U}_{tt'} \hat{Y}_{t't''i} + \hat{U}_{tt'} \hat{Z}_{t't''i} \right) \sum_{\tilde{v}_{t''i} \in \tilde{\mathcal{V}}_{t''i}} \lambda_{t''i}(\tilde{v}_{t''i}) \hat{v}_{t''i} \right] \\
& \quad \sum_{\tilde{v}_{t'i} \in \tilde{\mathcal{V}}_{t'i}} \lambda_{t'i}(\tilde{v}_{t'i}) = 1 \quad \forall t' = 1, \dots, t, i \in \hat{\mathcal{I}}_{t'} \\
& \quad \xi_{t'} \in \mathbb{R}^{K_{t'}} \quad \forall t' = 1, \dots, t \\
& \quad \left. \lambda_{t'i}(\tilde{v}_{t'i}) \in \mathbb{R}_+ \quad \forall t' = 1, \dots, t, i \in \hat{\mathcal{I}}_{t'}, \tilde{v}_{t'i} \in \tilde{\mathcal{V}}_{t'i} \right\} \\
= & \left\{ (\xi_{[t]}, \lambda_{[t]}) : \sum_{t'=1}^t \sum_{i \in \tilde{\mathcal{I}}_{t'}} w_{tt'i} \xi_{t'i} + \sum_{t'=1}^t \sum_{i \in \hat{\mathcal{I}}_{t'}} \sum_{\tilde{v}_{t'i} \in \tilde{\mathcal{V}}_{t'i}} \bar{w}_{tt'i} v_{t'i} \lambda_{t'i}(\tilde{v}_{t'i}) \right. \\
& \leq \sum_{t''=1}^{t-1} \sum_{i \in \hat{\mathcal{I}}_{t''}} \sum_{\tilde{v}_{t''i} \in \tilde{\mathcal{V}}_{t''i}} \sum_{t'=t''}^{t-1} \left[U_{tt'}^c \bar{X}_{t't''i}^c \tilde{v}_{t''i} \right. \\
& \quad \left. + \left(U_{tt'}^c \hat{X}_{t't''i}^c + U_{tt'}^b \hat{X}_{t't''i}^b + \bar{U}_{tt'} \hat{Y}_{t't''i} + \hat{U}_{tt'} \hat{Z}_{t't''i} \right) \hat{v}_{t''i} \right] \lambda_{t''i}(\tilde{v}_{t''i}) \\
& \quad \sum_{\tilde{v}_{t'i} \in \tilde{\mathcal{V}}_{t'i}} \lambda_{t'i}(\tilde{v}_{t'i}) = 1 \quad \forall t' = 1, \dots, t, i \in \hat{\mathcal{I}}_{t'} \\
& \quad \xi_{t'} \in \mathbb{R}^{K_{t'}} \quad \forall t' = 1, \dots, t \\
& \quad \left. \lambda_{t'i}(\tilde{v}_{t'i}) \in \mathbb{R}_+ \quad \forall t' = 1, \dots, t, i \in \hat{\mathcal{I}}_{t'}, \tilde{v}_{t'i} \in \tilde{\mathcal{V}}_{t'i} \right\}
\end{aligned}$$

Constraints in the multistage formulation (20) involving uncertain parameters are as follows:

$$\begin{aligned}
\sum_{t'=1}^t \sum_{t''=1}^{t'} \sum_{i \in \hat{\mathcal{I}}_{t''}} \left[\tilde{A}_{tt'}^c \bar{X}_{t't''i}^c \tilde{\zeta}_{t''i} + \left(\tilde{A}_{tt'}^c \hat{X}_{t't''i}^c + \tilde{A}_{tt'}^b \hat{X}_{t't''i}^b + \tilde{D}_{tt'} \hat{Y}_{t't''i} + \tilde{H}_{tt'} \hat{Z}_{t't''i} \right) \hat{\zeta}_{t''i} \right] & \leq \sum_{t'=1}^t \tilde{B}_{tt'} \xi_{t'} \\
\forall t \in \mathcal{T}, (\xi_{[t]}, \zeta_{[t]}, \tilde{\zeta}_{[t]}, \hat{\zeta}_{[t]}) & \in \tilde{\Theta}_t \left(\bar{X}_{[t-1]}^c, \hat{X}_{[t-1]}^c, \hat{X}_{[t-1]}^b, \hat{Y}_{[t-1]}, \hat{Z}_{[t-1]} \right),
\end{aligned}$$

which can similarly be reformulated using Λ_t as the uncertainty set for each $t \in \mathcal{T}$ and ξ and λ as

the uncertain parameters:

$$\begin{aligned}
& \sum_{t'=1}^t \sum_{t''=1}^{t'} \sum_{i \in \widehat{\mathcal{I}}_{t'}} \left[\tilde{\mathbf{A}}_{tt'}^c \overline{\mathbf{X}}_{t't''i}^c \sum_{\tilde{\mathbf{v}}_{t''i} \in \tilde{\mathcal{V}}_{t''i}} \lambda_{t''i}(\tilde{\mathbf{v}}_{t''i}) \tilde{\mathbf{v}}_{t''i} \right. \\
& \quad \left. + \left(\tilde{\mathbf{A}}_{tt'}^c \widehat{\mathbf{X}}_{t't''i}^c + \tilde{\mathbf{A}}_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b + \tilde{\mathbf{D}}_{tt'} \widehat{\mathbf{Y}}_{t't''i} + \tilde{\mathbf{H}}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \right) \sum_{\tilde{\mathbf{v}}_{t''i} \in \tilde{\mathcal{V}}_{t''i}} \lambda_{t''i}(\tilde{\mathbf{v}}_{t''i}) \hat{\mathbf{v}}_{t''i} \right] \leq \sum_{t'=1}^t \sum_{i \in \widehat{\mathcal{I}}_{t'}} \tilde{\mathbf{b}}_{tt'i} \xi_{t'i} \\
& \quad \forall t \in \mathcal{T}, (\boldsymbol{\xi}_{[t]}, \boldsymbol{\lambda}_{[t]}) \in \Lambda_t \left(\overline{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^b, \widehat{\mathbf{Y}}_{[t-1]}, \widehat{\mathbf{Z}}_{[t-1]} \right) \\
\Rightarrow & \sum_{t''=1}^t \sum_{i \in \widehat{\mathcal{I}}_{t''}} \sum_{\tilde{\mathbf{v}}_{t''i} \in \tilde{\mathcal{V}}_{t''i}} \sum_{t'=t''}^t \left[\tilde{\mathbf{A}}_{tt'}^c \overline{\mathbf{X}}_{t't''i}^c \tilde{\mathbf{v}}_{t''i} \right. \\
& \quad \left. + \left(\tilde{\mathbf{A}}_{tt'}^c \widehat{\mathbf{X}}_{t't''i}^c + \tilde{\mathbf{A}}_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b + \tilde{\mathbf{D}}_{tt'} \widehat{\mathbf{Y}}_{t't''i} + \tilde{\mathbf{H}}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \right) \hat{\mathbf{v}}_{t''i} \right] \lambda_{t''i}(\tilde{\mathbf{v}}_{t''i}) \leq \sum_{t'=1}^t \sum_{i \in \widehat{\mathcal{I}}_{t'}} \tilde{\mathbf{b}}_{tt'i} \xi_{t'i} \\
& \quad \forall t \in \mathcal{T}, (\boldsymbol{\xi}_{[t]}, \boldsymbol{\lambda}_{[t]}) \in \Lambda_t \left(\overline{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^c, \widehat{\mathbf{X}}_{[t-1]}^b, \widehat{\mathbf{Y}}_{[t-1]}, \widehat{\mathbf{Z}}_{[t-1]} \right).
\end{aligned}$$

We apply the worst case to each set of constraints associated with $t \in \mathcal{T}$:

$$\begin{aligned}
& \max_{(\boldsymbol{\xi}_{[t]}, \boldsymbol{\lambda}_{[t]}) \in \Lambda_t} \left\{ \sum_{t'=1}^t \sum_{i \in \widehat{\mathcal{I}}_{t'}} \sum_{\tilde{\mathbf{v}}_{t''i} \in \tilde{\mathcal{V}}_{t''i}} \sum_{t'=t''}^t \left[\tilde{\mathbf{A}}_{tt'}^c \overline{\mathbf{X}}_{t't''i}^c \tilde{\mathbf{v}}_{t''i} \right. \right. \\
& \quad \left. \left. + \left(\tilde{\mathbf{A}}_{tt'}^c \widehat{\mathbf{X}}_{t't''i}^c + \tilde{\mathbf{A}}_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b + \tilde{\mathbf{D}}_{tt'} \widehat{\mathbf{Y}}_{t't''i} + \tilde{\mathbf{H}}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \right) \hat{\mathbf{v}}_{t''i} \right] \lambda_{t''i}(\tilde{\mathbf{v}}_{t''i}) - \sum_{t'=1}^t \sum_{i \in \widehat{\mathcal{I}}_{t'}} \tilde{\mathbf{b}}_{tt'i} \xi_{t'i} \right\} \leq 0,
\end{aligned}$$

which can be formulated into the following set of constraints applying LP duality:

$$\begin{aligned}
& \sum_{t'=1}^t \sum_{i \in \widehat{\mathcal{I}}_{t'}} \psi_{tt'i} \leq \mathbf{0} \\
& \boldsymbol{\Phi}_t \mathbf{w}_{tt'i} = -\tilde{\mathbf{b}}_{tt'i} \quad \forall t' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t'} \\
& \psi_{tt'i} + \boldsymbol{\Phi}_t \left[\tilde{\mathbf{w}}_{tt'i} v_{t''i} - \sum_{t'=t''}^{t-1} \mathbf{U}_{tt'}^c \overline{\mathbf{X}}_{t't''i}^c \tilde{\mathbf{v}}_{t''i} \right. \\
& \quad \left. - \sum_{t'=t''}^{t-1} \left(\mathbf{U}_{tt'}^c \widehat{\mathbf{X}}_{t't''i}^c + \mathbf{U}_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b + \bar{\mathbf{U}}_{tt'} \widehat{\mathbf{Y}}_{t't''i} + \hat{\mathbf{U}}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \right) \hat{\mathbf{v}}_{t''i} \right] \\
& \geq \sum_{t'=t''}^t \left[\tilde{\mathbf{A}}_{tt'}^c \overline{\mathbf{X}}_{t't''i}^c \tilde{\mathbf{v}}_{t''i} + \left(\tilde{\mathbf{A}}_{tt'}^c \widehat{\mathbf{X}}_{t't''i}^c + \tilde{\mathbf{A}}_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b + \tilde{\mathbf{D}}_{tt'} \widehat{\mathbf{Y}}_{t't''i} + \tilde{\mathbf{H}}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \right) \hat{\mathbf{v}}_{t''i} \right]
\end{aligned}$$

$$\forall t \in \mathcal{T}, t'' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t''}, \tilde{\mathbf{v}}_{t''i} \in \tilde{\mathcal{V}}_{t''i}$$

$$\psi_{tt'i} \in \mathbb{R}^{P_t} \quad \forall t' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t'}$$

$$\boldsymbol{\Phi}_t \in \mathbb{R}_+^{P_t \times Q_t}.$$

This finally results in the reformulation (as presented in the paper):

$$\begin{aligned}
& \underset{\substack{\overline{\mathbf{X}}^c, \widehat{\mathbf{X}}^c, \widehat{\mathbf{X}}^b, \\ \widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}, \Phi, \Psi}}{\text{minimize}} & \mathbf{e}_1^\top \left(\overline{\mathbf{X}}_{111}^c + \widehat{\mathbf{X}}_{111}^c \right) \\
\text{subject to} & \sum_{t'=1}^t \sum_{i \in \widehat{\mathcal{I}}_{t'}} \psi_{tt'i} \leq \mathbf{0} \quad \forall t \in \mathcal{T} \\
& \Phi_t \mathbf{w}_{tt'i} = -\tilde{\mathbf{b}}_{tt'i} \quad \forall t \in \mathcal{T}, t' = 1, \dots, t, i \in \widetilde{\mathcal{I}}_{t'} \\
& \psi_{tt''i} + \Phi_t \left[\bar{\mathbf{w}}_{tt''i} v_{t''i} - \sum_{t'=t''}^{t-1} U_{tt'}^c \overline{\mathbf{X}}_{t't''i}^c \bar{\mathbf{v}}_{t''i} \right. \\
& \quad \left. - \sum_{t'=t''}^{t-1} \left(U_{tt'}^c \widehat{\mathbf{X}}_{t't''i}^c + U_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b + \bar{U}_{tt'} \widehat{\mathbf{Y}}_{t't''i} + \widehat{U}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \right) \hat{\mathbf{v}}_{t''i} \right] \\
& \geq \sum_{t'=t''}^t \left[\tilde{\mathbf{A}}_{tt'}^c \overline{\mathbf{X}}_{t't''i}^c \bar{\mathbf{v}}_{t''i} + \left(\tilde{\mathbf{A}}_{tt'}^c \widehat{\mathbf{X}}_{t't''i}^c + \tilde{\mathbf{A}}_{tt'}^b \widehat{\mathbf{X}}_{t't''i}^b + \tilde{\mathbf{D}}_{tt'} \widehat{\mathbf{Y}}_{t't''i} + \widetilde{\mathbf{H}}_{tt'} \widehat{\mathbf{Z}}_{t't''i} \right) \hat{\mathbf{v}}_{t''i} \right] \\
& \quad \forall t \in \mathcal{T}, t'' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t''}, \bar{\mathbf{v}}_{t''i} \in \widetilde{\mathcal{V}}_{t''i} \\
& \psi_{tt'i} \in \mathbb{R}^{P_t} \quad \forall t \in \mathcal{T}, t' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t'} \\
& \Phi_t \in \mathbb{R}_+^{P_t \times Q_t} \quad \forall t \in \mathcal{T} \\
& \left(\overline{\mathbf{X}}_{tt'i}^c, \widehat{\mathbf{X}}_{tt'i}^c, \widehat{\mathbf{X}}_{tt'i}^b, \widehat{\mathbf{Y}}_{tt'i}, \widehat{\mathbf{Z}}_{tt'i} \right) \in \Omega_{tt'i} \quad \forall t \in \mathcal{T}, t' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t'},
\end{aligned}$$

which can be further simplified into the following formulation if we assume $U_{tt}^c = U_{tt}^b = \bar{U}_{tt} = \widehat{U}_{tt} = \mathbf{0}$ for all $t \in \mathcal{T}$:

$$\begin{aligned}
& \underset{\substack{\overline{\mathbf{X}}^c, \widehat{\mathbf{X}}^c, \widehat{\mathbf{X}}^b, \\ \widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}, \Phi, \Psi}}{\text{minimize}} & \mathbf{e}_1^\top \left(\overline{\mathbf{X}}_{111}^c + \widehat{\mathbf{X}}_{111}^c \right) \\
\text{subject to} & \sum_{t'=1}^t \sum_{i \in \widehat{\mathcal{I}}_{t'}} \psi_{tt'i} \leq \mathbf{0} \quad \forall t \in \mathcal{T} \\
& \Phi_t \mathbf{w}_{tt'i} = -\tilde{\mathbf{b}}_{tt'i} \quad \forall t \in \mathcal{T}, t' = 1, \dots, t, i \in \widetilde{\mathcal{I}}_{t'} \\
& \psi_{tt''i} + \Phi_t \left[\bar{\mathbf{w}}_{tt''i} v_{t''i} - \sum_{t'=t''}^t \left(U_{tt'}^c + \tilde{\mathbf{A}}_{tt'}^c \right) \overline{\mathbf{X}}_{t't''i}^c \bar{\mathbf{v}}_{t''i} - \sum_{t'=t''}^t \left(\left[U_{tt'}^c + \tilde{\mathbf{A}}_{tt'}^c \right] \widehat{\mathbf{X}}_{t't''i}^c \right. \right. \\
& \quad \left. \left. + \left[U_{tt'}^b + \tilde{\mathbf{A}}_{tt'}^b \right] \widehat{\mathbf{X}}_{t't''i}^b + \left[\bar{U}_{tt'} \tilde{\mathbf{D}}_{tt'} \right] \widehat{\mathbf{Y}}_{t't''i} + \left[\widehat{U}_{tt'} + \widetilde{\mathbf{H}}_{tt'} \right] \widehat{\mathbf{Z}}_{t't''i} \right) \hat{\mathbf{v}}_{t''i} \right] \geq \mathbf{0} \\
& \quad \forall t \in \mathcal{T}, t'' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t''}, \bar{\mathbf{v}}_{t''i} \in \widetilde{\mathcal{V}}_{t''i} \\
& \psi_{tt'i} \in \mathbb{R}^{P_t} \quad \forall t \in \mathcal{T}, t' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t'} \\
& \Phi_t \in \mathbb{R}_+^{P_t \times Q_t} \quad \forall t \in \mathcal{T} \\
& \left(\overline{\mathbf{X}}_{tt'i}^c, \widehat{\mathbf{X}}_{tt'i}^c, \widehat{\mathbf{X}}_{tt'i}^b, \widehat{\mathbf{Y}}_{tt'i}, \widehat{\mathbf{Z}}_{tt'i} \right) \in \Omega_{tt'i} \quad \forall t \in \mathcal{T}, t' = 1, \dots, t, i \in \widehat{\mathcal{I}}_{t'}.
\end{aligned}$$

2 Data for Capacity Planning Case Study from Section 7.3

$r = 0.2$, $d_1^{\min} = 10,000$ kW, $d_1^{\max} = 15,000$ kW, $\bar{\delta} = 1.2$.

	α_i (k\$/kW)	β_i (k\$/kW)	C_{it} (kW)						
			Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	Period 7
Coal	2.89	0.00371	9,000	2,250	2,812	3,515	4,394	5,493	6,866
NG-GT	0.651	0.0299	7,500	1,875	2,344	2,930	3,662	4,577	5,722
Solar PV	3.480	0.0	1,500	375	468	586	732	915	1,145
Wind	1.980	0.0	1,500	375	468	586	732	915	1,145

	Period 2	Period 3	Period 4	Period 5	Period 6	Period 7
δ^{\min}	1.05	1.05	1.05	1.05	1.05	1.05
δ^{\max}	1.25	1.25	1.25	1.25	1.25	1.25