NORTHWESTERN UNIVERSITY

Optimization under Variable Uncertainty

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Industrial Engineering and Management Sciences

By

Kartikey Sharma

EVANSTON, ILLINOIS

March 2020

ABSTRACT

Optimization under Variable Uncertainty

Kartikey Sharma

In this dissertation, we study models and methods to address uncertainties that can vary in optimization problems. Robust optimization is a popular approach for optimization under uncertainty, especially if limited information is available about the distribution of the uncertainty. It models the uncertainty through sets and finds a robust optimal solution that is feasible for all realizations of the uncertainty within the set and is optimal for the worst-case realization. The structure of these sets determines the complexity of the resulting optimization problem. In most models, the uncertainty set is assumed to be *exogenous* i.e., pre-determined and is unaffected by decisions or other uncertainty realizations in the problem. This thesis introduces *endogenous* uncertainty models, which may be affected by decisions that are made in the problem or by other uncertainty realizations within the problem.

In the first chapter, we take a step towards generalizing robust linear optimization to problems with *decision dependent uncertainties*. We show these problems to be NP-complete in general settings. To alleviate these computational inefficiencies, we introduce

a class of uncertainty sets whose sizes depend on binary decisions. We propose reformulations that improve upon alternative standard linearization techniques. To illustrate the advantages of this framework, a shortest path problem is discussed, where the uncertain arc lengths are affected by decisions. The proposed notion of proactive uncertainty control provides modeling and performance advantages, and mitigates over conservatism of common robust optimization approaches.

While the impact of the decisions on the uncertainty set was fixed in the first chapter, we extend the decision-dependent models to allow for uncertainty in the influence of decisions on the sets in the second chapter. Here, the exact impact of the decision on the uncertainty set itself may be uncertain. This situation arises in many practical settings where the decision's impact may not be known a priori. It is especially relevant for problems in which the decision is on the gathering of information. We leverage robust and stochastic optimization to incorporate uncertain influence into the optimization problem. We then evaluate the performance of these models on a power systems unit commitment problem.

The third chapter discusses the topic of *Connected Uncertainties*, i.e., uncertainty models in which past realizations influence future uncertainties. For this class of problems, we develop a novel modeling framework that naturally incorporates this dependence via connected uncertainty sets, whose parameters at each period depend on previous uncertainty realizations. To find optimal here-and-now solutions, we reformulate robust and distributionally robust constraints for popular set structures and demonstrate this modeling framework numerically on broadly applicable knapsack and portfolio optimization problems.

In the fourth chapter of the thesis, we leverage the idea of connected uncertainty to develop *robust adaptive classifiers* for streaming data. Classification algorithms are effective, when data can be modeled by time-invariant distributions. In streaming settings, a classifier needs to be updated continuously, and hence static classifiers lose their reliability over time. We consider streaming data sets in which the behavior of each class can be modeled by a time series. For classification of such streaming data, we extend the Minimax Probability Machine to incorporate a time series model using the principles of connected uncertainty sets. We illustrate the new methods by numerical experiments on synthetic data.

Overall, this thesis led to insights in two directions. First, we introduced uncertainty sets which depended on decisions. This enabled us to model reducing the uncertainty at a price, which is common in practical applications. This approach also allowed us to capture many problems in which the uncertainty naturally depends on decisions. In the second direction, we studied multi-period problems where today's uncertainty can affect the uncertainty tomorrow. This led us to capture correlations over time, which are common in many applications. Our future goal is to further extend this work in both directions. Specifically, we want to solve larger unit commitment problems, solve the problem of continuous variables affecting uncertainty sets and merge decision dependent and connected uncertainties.

Acknowledgements

I would like to express my gratitude to all the people who made my PhD memorable, and I am grateful to be able to acknowledge them here.

First and foremost, my deepest gratitude goes to my advisor, Prof. Omid Nohadani for his guidance during my PhD. It is hard to imagine this thesis without his support, patience and encouragement.

I am indebted to Prof. David Morton for his support on my research projects, as well as serving on my committee. I would also like to thank Prof. John Birge and Prof. Daniel Kuhn for serving on my PhD committee.

The past five years spent at the IEMS department have been a true pleasure thanks to my friends, the professors and the staff of the department. My heartfelt thanks go out to them. I would like to give special thanks to my roommate Raghu for being there for me over the years.

Finally, I would like to thank my parents and the rest of my family without whose continuous love and support this thesis would not have been possible.

Dedication

To my parents, Anil and Meenakshi

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Optimization problems are a corner stone of modern life. They occur in many applications and solving them is key to improving the performance of various systems. General optimization problems are of the form,

$$\begin{aligned} &\min_{\mathbf{x}} f(\mathbf{x}) \\ &\text{s.t. } g_i(\mathbf{x}) \geq 0 \quad \forall i \in \mathcal{I} \\ &\mathbf{x} \in \mathcal{X}. \end{aligned} \tag{1.1}$$

Here, the goal is to find a decision $\mathbf{x}$ that lies in the set $\mathcal{X}$, satisfies the constraints $g_i(\mathbf{x}) \geq 0 \ \forall i \in \mathcal{I}$, and minimizes the objective. In practical applications, the set $\mathcal{X}$ and the constraints $g_i(\mathbf{x}) \geq 0$, represent physical limitations, budgetary constraints, or model assumptions etc. The objective function $f(\mathbf{x})$ represents factors such as cost, utility, flows etc. These functions $f(\mathbf{x})$ and $g_i(\mathbf{x})$ are constructed from real systems and as such use various parameters and their estimates. They may also use some parameters whose values are inherently uncertain and cannot be known a priori. These sources of uncertainty limit the scope of the problem (1.1), as any solution arising from this problem depends on the parameter estimates used. If the true value, or the realization of the parameters, is different from the estimate, it may lead to suboptimal or even infeasible solutions. This danger is particularly relevant for optimal solutions since they tend to exist on boundaries.

Two well-established approaches for optimization problems under such uncertainty are stochastic and robust optimization.

Stochastic Optimization

Stochastic optimization (SO) assumes that the source of uncertainty can be described by a distribution that is known to the decision maker. This distribution is leveraged to solve the optimization problem. A general form of a SO problem is,

$$\begin{aligned} &\min_{\mathbf{x}} \mathbb{E}[f(\mathbf{x}, \boldsymbol{\xi})] \\ &\text{s.t. } \mathbb{P}[g_i(\mathbf{x}, \boldsymbol{\xi}) \geq 0 \quad \forall i \in \mathcal{I}] \geq 1 - \alpha \\ &\mathbf{x} \in \mathcal{X}. \end{aligned} \tag{1.2}$$

Here, we minimize the expected value of the objective function while trying to satisfy the constraints with probability $1 - \alpha$. SO problems are primarily solved using techniques which leverage the structure of the optimization problem. These methods include Sample Average Approximation [87] to approximate the expectation and the chance constraints, cut generation methods for multistage problems [34], approximation techniques [26] for evaluating solutions etc. These approaches leverage the structure of the optimization problems to make up for the increase in problem size due to the existence of scenarios.

Robust Optimization

A key requirement for SO problems is the knowledge of the distribution which may be unavailable in some settings. Robust optimization (RO) is a computationally attractive alternative [12, 22] to SO, where we only require knowledge of the set, in which the

uncertainty lies. RO problems can be expressed as following

$$\begin{aligned} &\min_{\mathbf{x}} \max_{\boldsymbol{\xi}\in\mathcal{U}} [f(\mathbf{x},\boldsymbol{\xi})] \\ &\text{s.t. } g_i(\mathbf{x},\boldsymbol{\xi}) \geq 0 \quad \forall \boldsymbol{\xi} \in \mathcal{U},\ \forall i \in \mathcal{I} \\ &\mathbf{x} \in \mathcal{X}. \end{aligned} \tag{1.3}$$

In problem (1.3), the uncertainty is assumed to lie in a set $\mathcal{U}$. We optimize for the worst-case objective over all $\boldsymbol{\xi}$ in this set. Any solution $\mathbf{x}$ is required to be feasible for all possible realizations of the uncertain parameter in $\mathcal{U}$. The method of RO has been developed considerably and applied to problems ranging from portfolio management [45], to healthcare [30], to electricity systems [74], and to engineering design [21].

The geometry of the set $\mathcal{U}$ also determines the computational tractability of the problem. For example, certain combinatorial RO problems achieve a tractable reformulation when the uncertain objective coefficients reside in a cardinality constrained set [3, 18]. The size of such sets controls the magnitude of possible uncertainties, to which the solution is immune. It also establishes probabilistic guarantees of constraint satisfaction.

Distributionally Robust Optimization

When the adversarial uncertain components exhibit probabilistic characteristics, distributionally robust optimization (DRO) offers an alternative approach by replacing standard

uncertainty sets with ambiguity sets over distributions [35, 48],

$$\begin{aligned} &\min_{\mathbf{x}} \max_{\mathbb{P}\in\mathcal{U}} \mathbb{E}[f(\mathbf{x},\boldsymbol{\xi})] \\ &\text{s.t. } \mathbb{P}[g_i(\mathbf{x},\boldsymbol{\xi}) \geq 0 \quad \forall i \in \mathcal{I}] \geq 1-\alpha \quad \forall \mathbb{P} \in \mathcal{U}, \\ &\mathbf{x} \in \mathcal{X}. \end{aligned} \tag{1.4}$$

Here, we consider the worst-case expectation over all distributions lying in the set $\mathcal{U}$. These sets can be characterized by moments [99], by distance measures [13, 39, 43], or by hypothesis tests [23]. DRO techniques have been applied to a broad range of applications, such as portfolio management [79], simulation [66], and supply chain [42] problems.

Adjustable Robust Optimization

Most RO approaches focus on static *here-and-now* solutions. However, the benefits of these solutions are limited in many settings, especially those that can accommodate adaptation. The extension of RO to multistage problems [11] has revealed the deficiencies of static *here-and-now* solutions. Since these solutions do not adapt to uncertainty realizations, they lead to highly conservative solutions. *Wait-and-see* decisions adapt to uncertainty realizations like recourse in SO. Such an adjustable robust optimization (ARO) problem can be expressed as

$$\begin{aligned} &\min_{\mathbf{x}} f(\mathbf{x}) + \max_{\boldsymbol{\xi}\in\mathcal{U}} \min_{\mathbf{y}\in\mathcal{Y}(\mathbf{x},\mathbf{y},\boldsymbol{\xi})} h(\mathbf{x},\mathbf{y},\boldsymbol{\xi}) \\ &\text{s.t. } \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{1.5}$$

In this problem, the second-stage decision $\mathbf{y}$ can adapt to the uncertainty realization $\boldsymbol{\xi}$. This approach improves the solution quality at the expense of higher computational

complexity. Decision rules can provide a smooth trade-off between complexity and solution quality [62] for ARO problems, by limiting the flexibility of the second-stage decision $\mathbf{y}$. This makes solving the problem easier but leads to suboptimal solutions to the original problem. In the context of multistage DRO, solutions adapt to the realization of the uncertainty instead of adapting to the realized distribution. Non-anticipative decision rules can be leveraged to provide tractable reformulations for moment-based uncertainty sets [38, 48]. Furthermore, adaptability has been extended to ambiguity sets defined by the Wasserstein metric with a conic reformulation for a two-stage DRO problem [50].

Uncertainty Models

RO and DRO problems, both use sets to model the uncertainty. As mentioned earlier, the geometry of the uncertainty set determines the computational tractability of the robust problem as well as the amount of protection provided against the uncertainty. Some commonly used standard uncertainty sets for RO problems are

- Elliposidal Uncertainty set $\{\boldsymbol{\xi} \mid \boldsymbol{\xi} = \boldsymbol{\mu} + \mathbf{L}\mathbf{u},\ \|\mathbf{u}\|_2 \leq r\}$
- Polyhedral Uncertainty set $\{\boldsymbol{\xi} \mid \mathbf{D}\boldsymbol{\xi} \leq \mathbf{d},\ \boldsymbol{\xi} \geq \mathbf{0}\}$
- Cardinality Constrained Uncertainty set

$$\mathcal{U} = \left\{ \boldsymbol{\xi} \;\middle|\; \sum_{i=1}^{N} \frac{|\xi_i - \bar{\xi}_i|}{\hat{\xi}_i} \leq \Gamma\sqrt{N},\ \ \xi_i \in [\bar{\xi}_i - \Gamma\hat{\xi}_i, \bar{\xi}_i + \Gamma\hat{\xi}_i] \right\}.$$

Depending on the constraint being reformulated, the uncertainty set used, and the true nature of the uncertainty, it is possible to achieve probabilistic guarantees on the constraint satisfaction of the robust solution [18]. For DRO problems, the uncertainty sets

are over distributions instead of uncertainty realizations. These sets are described by moments or metrics on the space of distributions. For example

- Moment Uncertainty set $\{P \in \mathcal{M} \mid \int_{\Xi} dP(\boldsymbol{\xi}) = 1,\ \int_{\Xi} \boldsymbol{\xi} dP(\boldsymbol{\xi}) = \boldsymbol{\mu}\}$, where $\mathcal{M}$ is the set of finite measures
- Ball uncertainty set $\{P \in \mathcal{M} \mid d(P, P_N) \leq r\}$ where P_N is a nominal distribution and $d(\cdot, \cdot)$ is a metric on the space of distributions.

For both static and adaptive decisions, RO models primarily use pre-determined uncertainty sets. These sets are unaffected by the decisions being made in the optimization problem or by other uncertainty realizations in the problem. Our goal in this dissertation is to address RO problems where the uncertainty set can be variable and affected by either the decisions in the problem or other uncertainty realizations.

Notation

Throughout this dissertation, we use bold lower and uppercase letters to denote vectors and matrices. Scalars are marked in regular font. All vectors are column vectors and the vector of ones is denoted by $\mathbf{e}$. For any given matrix $\mathbf{A}$, the i^{th} row is denoted by $\mathbf{A}_{i,\bullet}$ and the j^{th} column is denoted by $\mathbf{A}_{\bullet,j}$. Furthermore, $diag(\bullet)$ denotes a diagonal matrix with $\bullet$ on the diagonal and zeros elsewhere. LHS denotes left-hand-side and RHS denotes right-hand-side. We use the phrases "decision-dependent" and "endogenous" interchangeably. Similarly, we refer to variables affecting an uncertainty set as influence variables. To streamline the exposition, we use "uncertainty set" for both the RO and DRO settings. For the former, the set is over parameters and for the latter it is over distributions and is

also known as an ambiguity set. T denotes the total time periods, τ refers to a particular time period, t serves as an index, and $\mathbf{e}$ is a vector of all ones.

CHAPTER 2

Decision Dependent Uncertainty

2.1. Introduction

RO employs uncertainty sets that are predetermined and, hence, *exogenous*. However, in many real-world problems, the uncertainty can be affected by decisions. In such decision-dependent cases, the uncertainty set is *endogenous*. Despite the wide prevalence of such uncertainties in real-world settings, these problems have not received much attention in the past, largely due to computational intractabilities. In this chapter, we take a first step towards robust linear optimization problems with endogenous uncertainties and provide a class of uncertainty sets, whose reformulations improve over standard techniques. Specifically, we study a single-stage RO problem with decision-dependent uncertainty sets

$$\text{(RO-DDU)} \qquad \begin{aligned} &\min_{\mathbf{x},\mathbf{y}} \ \mathbf{c}^\top \mathbf{x} + \mathbf{f}^\top \mathbf{y} \\ &\text{s.t. } \mathbf{a}_i^\top \mathbf{x} + \boldsymbol{\xi}_i^\top \mathbf{y} \leq b_i \quad \forall \boldsymbol{\xi}_i \in \mathcal{U}_i(\mathbf{x}) \subseteq \mathbb{R}^n \quad \forall i = 1, \ldots, m, \end{aligned}$$

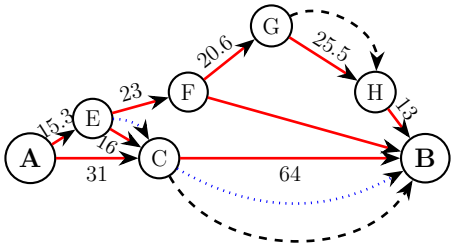
where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$ represent decision variables, which need to satisfy each constraint $i = 1, \ldots, m$ for every realization from the set $\mathcal{U}_i(\mathbf{x})$ that bounds the uncertain parameter $\boldsymbol{\xi}_i$. Further, the parameters defining $\mathcal{U}_i(\mathbf{x})$ depend on decisions $\mathbf{x}$. We first study the complexity of (RO-DDU) for polyhedral $\mathcal{U}_i(\mathbf{x})$. We then assume $\mathbf{x}$ is binary

and provide reformulations for special classes of polyhedral and conic uncertainty sets and conclude with numerical experiments.

To show the range of applicability of this model, we illustrate two examples.

Example 1: Uncertainty Reduction. In facility location or inventory management problems with uncertain demand, the uncertainty can be reduced by spending resources to acquire information. Similarly, in healthcare problems, additional medical tests can improve the diagnosis. This type of uncertainty reduction is characteristic of many real-world problems. In order to improve solutions, decisions on uncertainty reduction have to be included into the optimization problem, making the uncertainty a function of decisions on acquiring additional information.

Example 2: Shortest Path on a Network. Consider the graph in Figure 2.1 with the arcset $\mathcal{A}$ and let the uncertain length for any arc e be $d_e = \bar{d}_e(1 + 0.5\xi_e)$, where $\bar{d}_e$ denotes the nominal value. The uncertain vector $\boldsymbol{\xi}$ lies in the uncertainty set $\mathcal{U}(\mathbf{x}) = \left\{\boldsymbol{\xi} \mid 0 \leq \xi_e \leq 1 - 0.8x_e \ \ \forall e, \ \sum_{e\in\mathcal{A}} \xi_e \leq 1\right\}$. The binary decision x_e determines whether to reduce the maximum possible uncertainty ξ_e to 0.2 ($x_e = 1$) or leave it at 1 ($x_e = 0$). For simplicity, we assume the reduction to be possible for at most one of the arcs.



Shortest Path	Path	Nominal	Worst-case
Nominal	A−C−B	95	$31 + 1.5 \times 64 =$ **127**
Robust	A−E−F−G−H−B	97.4	$15.3 + 23 + 20.6 +$ $1.5 \times 25.5 + 13$ $=$ **110.15**
Endogenous Robust	A−E−C−B	95.3	$15.3 + 1.4 \times 16 +$ $1.1 \times 64 =$ **108.1**

Figure 2.1. Shortest path on a network. Nominal lengths are labeled. Worst-case and reduced-case lengths are displayed with dashed and dotted lines. The table shows the lengths in different settings.

Figure 2.1 displays a network with source node A and destination B. In the worst-case, the nominally shortest path lengthens to 127 units. RO optimizes against this case, improving the worst-case length. If it is permitted to reduce the uncertainty of an arc, then A−E−C−B is selected with $x_{C-B} = 1$ and the worst-case path becomes 108.5. This example demonstrates that decision-dependent sets can be leveraged to model decisions that mitigate the worst-case scenario.

The contributions of this chapter can be summarized as follows:

(1) We study robust linear optimization problems with a polyhedral decisiondependent set for the uncertain parameters. We prove such problems to be NP-complete. We also show that when decisions that influence the uncertainties are binary, the problem can be reformulated as a mixed integer optimization problem.

(2) For binary $\mathbf{x}$, we provide a class of uncertainty sets for which a more efficient reformulation of the decision-dependent RO problem is possible. The set structure and the nature of decision dependence are leveraged to provide reformulations with fewer constraints.

(3) We provide an improvement to Big-M linearization for bilinear terms which can reduce the number of constraints.

This chapter also showcases the advantages that can be gained in both stochastic and robust optimization by proactively controlling uncertainties.

We also emphasize what this chapter fails to address. Reformulations for continuous decisions influencing the uncertainty are not provided. Furthermore, the primary problem

in this chapter is a static optimization problem, i.e., the decisions do not adapt to uncertainty realizations. In fact, it is the uncertainty set and the corresponding worst-case realization that are affected by decisions.

Section 2.3 discusses the complexity of the decision-dependent robust linear optimization problem. Section 2.4 introduces a class of uncertainty sets which allow improved reformulations. Section 2.5 provides a comparison to the corresponding Big-M formulation. It also provides methods to improve these standard techniques. A numerical experiment is discussed in Section 2.6 to illustrate the advantages of the decision-dependent setting and to computationally compare the three formulations.

2.2. Background

In the following, we first review endogenous settings in SO before discussing RO approaches.

The notion of endogenous uncertainty in SO generally corresponds to scenario trees, where decisions determine the probabilities. For example, Jonsbråten et al. [58] consider the cost of an item to remain uncertain until it is produced. The probability distribution depends upon which item is to be produced and when. Goel and Grossmann [46] address the problem of offshore oil and gas planning, with the objective of maximizing revenues and investments over a period of time, when the recovery and size of oil fields are not known in advance. They provide a disjunctive formulation that is solved by a decomposition algorithm. This approach is extended to a multistage SO problem for optimal production scheduling, that minimizes cost while satisfying the demand for different goods [47]. For package sorting centers, Novoa et al. [83] seek to balance the flow across

working stations. Capacities are modeled via budgeted uncertainties where the budget is a function of workstation allocation. These and other approaches address endogenous uncertainties probabilistically.

In RO, the endogenous nature of uncertainty is imposed directly on the uncertainty set itself. For example, Spacey et al. [91] address a software partitioning problem, where code segments are assigned to different computing nodes to reduce runtime with uncertain execution order and for unknown frequency of segment calls. They employ tailored decision-dependent uncertainty sets. Such sets also occur as a result of reformulations. For example, Hanasusanto et al. [51] use a finite adaptability approximation to adjustable robust optimization (ARO), as introduced by Bertsimas and Caramanis [16], and consider optimization problems with binary recourse decisions. For problems with uncertain objective and constraints, they provide a formulation with decision-dependent uncertainty sets before finally reformulating it as a MILP. Poss [84, 85] considers combinatorial optimization problems with budgeted uncertainty sets. This extends the work of Bertsimas and Sim [18] to decision-dependent budgets. These works focus on budget uncertainty sets with limited discussion on general sets. On the other hand, for a dynamic pricing problem with learning, Bertsimas and Vayanos [20] consider 1 or ∞-norm uncertainty sets for price-dependent demand. Specifically, the uncertain demand curve is driven by past realizations of price-demand pairs. Since the price is a decision variable, this leads to decision-dependent uncertainty sets. In the context of robust scheduling problems, Vujanic et al. [96] consider a decision-dependent uncertainty set which is a vector sum of a collection of sets. The sets in the vector combination are selected by a decision which is a part of the original problem. They probe the performance of an affine policy for the

problem. More recently, decision-dependent sets were studied in the context of control problems with *primitive* uncertainty sets [104]. Note that in all approaches to date, the decision dependence is modeled in a specific context, often driven by an application.

The journey of RO has also included measures to reduce conservatism. The original RO formulation by Soyster [90] produced over conservative solutions for many applications due to the use of box uncertainties. Later, Ben-Tal and Nemirovski [10] provided less conservative solutions by using general polyhedral and ellipsoid uncertainty sets. ARO models [11] and decision rule approximations took another step in this direction by allowing decisions to depend on the realizations [44, 55]. In this vein, decision-dependent uncertainty sets offer a new avenue to reduce the level of conservatism. For example, Poss [84] decreases it for cardinality constrained sets. This chapter also motivates the notion of *proactive uncertainty control* by using decision-dependent sets to enable deliberate uncertainty reduction.

2.3. General Decision Dependence

Robust linear optimization problems encompass a wide variety of applications, in portfolio optimization, healthcare, inventory management, and routing, amongst others. The tractability of robust linear programs provides a suitable starting point to analyze the complexity of RO problems with decision-dependent uncertainty. Here, we investigate a robust linear optimization problem as in (RO-DDU). The underlying uncertainty set is endogenous and defined as follows.

Definition 2.3.1. *The set with constraint matrix* $\mathbf{D}$*, constant vector* $\mathbf{d}$*, and decision coefficient matrix* $\mathbf{\Delta}$ *given by*

$$\mathcal{U}^P(\mathbf{x}) = \{\boldsymbol{\xi} \mid \mathbf{D}\boldsymbol{\xi} \leq \mathbf{d} + \mathbf{\Delta}\mathbf{x}\}$$

is a ***polyhedral*** *uncertainty set with affine decision dependence.*

Note that $\mathbf{\Delta}$ determines the influence of $\mathbf{x}$ on the set and can be estimated from the data or from the context of an application. In Section 2.6, we quantify it for a specific application.

The following theorem shows that RO problems with decision-dependent sets cannot be reformulated in a tractable fashion, a departure from standard RO problems. This occurs despite the fact that linear programs with polyhedral uncertainty sets have tractable robust counterparts.

Theorem 2.3.1. *The robust linear problem* (RO-DDU) *with uncertainty set* $\mathcal{U}^P$ *is NP-complete.*

Proof. The proof follows the following steps:

(1) Consider an instance of the 3-Satisfiability problem (3-SAT) for a set of literals $N = \{1, 2, \ldots, n\}$ and m clauses, which seeks to find a solution $\mathbf{x} \in \{0, 1\}^n$ that satisfies

$$x_{i_1} + x_{i_2} + (1 - x_{i_3}) \geq 1 \text{ for } m \text{ clauses and } i_1, i_2, i_3 \in \{1, \ldots, n\}.$$

(2) Consider the following special case of (RO-DDU) with $\mathbf{x} \in \mathbb{R}^n$ $\mathbf{y} \in \mathbb{R}^m$, $z \in \mathbb{R}$

$$\text{(RO-SAT)} \qquad \min_{\mathbf{x},\mathbf{y},z\geq 0} \left\{ -z \mid z - \boldsymbol{\xi}^\top \mathbf{y} \leq 0, \ \ \forall \boldsymbol{\xi} \in \mathcal{U}(\mathbf{x}), \ \ \mathbf{x}, \mathbf{y} \leq \mathbf{e}, \ \ -\mathbf{y} \leq -\mathbf{e} \right\},$$

where $\mathcal{U}(\mathbf{x}) = \{(\xi_1, \ldots, \xi_m) \mid \xi_i \geq x_{i_1} \ , \ \xi_i \geq x_{i_2} \ , \ \xi_i \geq 1 - x_{i_3} \ , \ \xi_i \leq 1\}$.

Note that the 3-SAT problem is embedded in this set.

(3) By Lemma 2.3.1 (provided after these steps), the optimal value of (RO-SAT) is $-m$, if and only if the 3-SAT problem has a solution.

(4) Problem (RO-SAT) is a special case of (RO-DDU) with polyhedral set $\mathcal{U}(\mathbf{x})$.

(5) Since the 3-SAT problem is NP-complete [31], problem (RO-DDU) is also NP-complete.

□

Lemma 2.3.1. *The 3-SAT problem has a feasible solution* $\mathbf{x}$*, if and only if problem* (RO-SAT) *has an optimal value of at most* $-m$*.*

Proof. ($\Longrightarrow$) Suppose the 3-SAT problem has a feasible solution $\mathbf{x}$. This means, $\mathbf{x}$ has to satisfy

$$x_{i_1} + x_{i_2} + (1 - x_{i_3}) \geq 1 \ \ \forall i = 1, \ldots, m.$$

Since $\mathbf{x} \in \{0,1\}^n$, for each i at least one of x_{i_1}, x_{i_2}, or $1 - x_{i_3}$ must be equal to 1. Now, consider the uncertainty set

$$\mathcal{U}(\mathbf{x}) = \{(\xi_1, \ldots \xi_m) \mid \xi_i \geq x_{i_1} \ , \ \xi_i \geq x_{i_2} \ , \ \xi_i \geq 1 - x_{i_3} \ , \ \xi_i \leq 1 \ \ \forall i = 1, \ldots, m\}.$$

Since at least one of x_{i_1}, x_{i_2}, or $1 - x_{i_3}$ equals 1, ξ_i satisfies $\xi_i \geq 1$. This implies that $\xi_i = 1 \ \forall i$ is the only point in $\mathcal{U}(\mathbf{x})$. For this uncertainty set, the feasible solution is $\mathbf{x}, \mathbf{y} = \mathbf{1}, \ z = m$. This leads to the optimal solution $-z \leq -m$ or $z \geq m$.

($\Longleftarrow$) Suppose (RO-SAT) has an optimal solution $(\mathbf{x}^*, \mathbf{y}^*)$ with the objective value of $-z^* \leq -m$. We first show that strict inequality is not possible. Assume $-z^* < -m$. The constraints in (RO-SAT) imply $z^* - \boldsymbol{\xi}^\top \mathbf{y}^* \leq 0$, i.e., $\boldsymbol{\xi}^\top \mathbf{y}^* \geq z^* > m \ \forall \boldsymbol{\xi} \in \mathcal{U}(\mathbf{x}^*)$. The constraints also imply $y_i^* = 1 \ \forall i$. This means that $\sum_{i=1}^m \xi_i > m \ \forall \boldsymbol{\xi} \in \mathcal{U}(\mathbf{x}^*)$. However, the construction of the uncertainty set yields $\xi_i \leq 1$. This leads to a contradiction, because $\sum_{i=1}^m \xi_i \not> m$, and hence $-z^* = -m$. Thus, $\boldsymbol{\xi}^\top \mathbf{y}^* = m \ \forall \boldsymbol{\xi} \in \mathcal{U}(\mathbf{x}^*)$. Therefore, we can write $\sum_{i=1}^m \xi_i = m \ \ \forall \boldsymbol{\xi} \in \mathcal{U}(\mathbf{x}^*)$, which implies $\min_{\boldsymbol{\xi} \in \mathcal{U}(\mathbf{x}^*)} \sum_{i=1}^m \xi_i = m$. However, since the uncertainty set implies $\xi_i \leq 1 \ \forall i$, we can conclude that the sum can only be equal to m, if $\xi_i = 1 \ \ \forall i$.

We now show that this result implies for each i at least one of $x_{i_1}^*$ or $x_{i_2}^*$ or $(1 - x_{i_3}^*)$ is equal to 1. Suppose this is not true. This implies $\exists i$ for which $x_{i_1}^* < 1$, $x_{i_2}^* < 1$ and $(1 - x_{i_3}^*) < 1$. That means that we can construct $\xi_i' = \max\{x_{i_1}^*, x_{i_2}^*, (1 - x_{i_3}^*)\}$ which is an element of the uncertainty set and $\xi_i' < 1$. However, this contradicts the result of $\xi_i = 1 \ \ \forall i$. Therefore, if $z^* = m$, then we can find a feasible solution for the 3-SAT problem. □

Although problem (RO-DDU) is NP-complete, it can be reformulated as a bilinear or biconvex program, which may be solved by global optimization techniques [e.g., 60]. For binary decision variables $\mathbf{x}$ influencing $\mathcal{U}(\mathbf{x})$, the problem (RO-DDU) can be reformulated as a MILP, using the Big-M method (see Section 2.5). However, they suffer from weak relaxations.

2.4. Structured Uncertainty Sets

The weak numerical performance of Big-M linearization can be overcome, if the decision $\mathbf{x}$ plays a decisive role in governing the elements of the uncertainty set. Specifically, if the effect of $\mathbf{x}$ on the uncertainty set constraints can be modeled by penalizing the objective coefficients, then the number of constraints in the robust counterpart can be reduced. Here, we discuss the setting where $\mathbf{x}$ controls the upper bounds of the uncertain variables. This mechanism can be expressed in the set:

$$\textbf{$\overline{\mathbf{\Pi}}$-Uncertainty:} \quad \mathcal{U}^{\overline{\mathbf{\Pi}}}(\mathbf{x}) = \{\boldsymbol{\xi} \mid \mathbf{D}\boldsymbol{\xi} \leq \mathbf{d},\ \ \boldsymbol{\xi} \leq \mathbf{v} + \mathbf{W}(\mathbf{e} - \mathbf{x}),\ \ \boldsymbol{\xi} \geq \mathbf{0}\}\,.$$

Here, $\mathbf{D} \in \mathbb{R}^{m \times n}$ is a coefficient matrix, $\mathbf{d} \in \mathbb{R}^m$ is the RHS vector, $\mathbf{v} \in \mathbb{R}^n_+$ are the minimum upper bounds, and $\mathbf{W} = diag(\mathbf{w}) \in \mathbb{R}^{n \times n}_+$ (a diagonal matrix) are the incremental upper bounds, which apply when reduction is not applied. For $\mathcal{U}^{\overline{\mathbf{\Pi}}}$, the influence variable is $\mathbf{x} \in \{0,1\}^n$. The decision dependence in $\mathcal{U}^{\overline{\mathbf{\Pi}}}$ affects the upper bounds on each uncertain component ξ_i. This means, if the problem allows influencing uncertainties, this set can model *proactive* uncertainty reduction. One possible example is disaster planning, where a decision to reduce the fragility of certain roads yields an improved worst-case outcome. Another example is measurement applications, where a decision for additional expenditure leads to increased accuracy. We employed such a set in Example 2 and discuss it further in the numerical application.

We now discuss how this structure can be leveraged to reformulate the original problem (RO-DDU). Note that the objective function remains unaffected by the definition of the uncertainty set, as does the first term of the constraint. Therefore, we focus our

discussion on the parts of the constraint in problem (RO-DDU), that are affected by uncertainty.

2.4.1. $\overline{\mathbf{\Pi}}$-Uncertainty

For succinctness, this section provides a reformulation of the following linear constraint

$$\text{(LC)} \qquad \mathbf{y}^\top \boldsymbol{\xi} \leq b \;\; \forall \boldsymbol{\xi} \in \mathcal{U}^{\overline{\mathbf{\Pi}}}(\mathbf{x}).$$

To satisfy this constraint for all $\boldsymbol{\xi} \in \mathcal{U}^{\overline{\mathbf{\Pi}}}(\mathbf{x})$, the uncertain LHS needs to be replaced by its maximum over the set. For this, consider the following two problems:

$$\text{(P)} \qquad \begin{aligned} h(\mathbf{x},\mathbf{y}) = \\ \max_{\boldsymbol{\xi}} \;\; & \mathbf{y}^\top \boldsymbol{\xi} \\ \text{s.t.} \;\; & \mathbf{D}\boldsymbol{\xi} \leq \mathbf{d} \\ & \boldsymbol{\xi} \leq \mathbf{v} + \mathbf{W}(\mathbf{e} - \mathbf{x}) \quad : \boldsymbol{\pi}(\mathbf{x},\mathbf{y}) \\ & \boldsymbol{\xi} \geq \mathbf{0}, \end{aligned}$$

$$\text{(P')} \qquad \begin{aligned} \bar{h}(\mathbf{x},\mathbf{y}) = \\ \max_{\boldsymbol{\xi},\boldsymbol{\zeta}} \;\; & (\mathbf{y} - \overline{\mathbf{\Pi}}\mathbf{x})^\top \boldsymbol{\xi} + \mathbf{y}^\top \boldsymbol{\zeta} \\ \text{s.t.} \;\; & \mathbf{D}\boldsymbol{\xi} + \mathbf{D}\boldsymbol{\zeta} \leq \mathbf{d} \\ & \boldsymbol{\xi} \leq \mathbf{W}\mathbf{e} \\ & \boldsymbol{\zeta} \leq \mathbf{v} \\ & \boldsymbol{\xi}, \boldsymbol{\zeta} \geq \mathbf{0}, \end{aligned}$$

where in problem (P), $\boldsymbol{\pi}(\mathbf{x},\mathbf{y})$ denotes the corresponding dual variable. Problem (P) maximizes the LHS directly over $\mathcal{U}^{\overline{\mathbf{\Pi}}}(\mathbf{x})$. However, the standard reformulation of this problem leads to bilinear terms. To avoid them, we can leverage the structure of the uncertainty set and formulate problem (P) as problem (P'). Such a problem pair was also suggested in the context of stochastic network interdiction [32]. Proposition 2.4.1 uses the duals of (P) and (P') to prove that they have the same objective value at optimality. Formulating

problem (P') requires the use of matrix $\overline{\mathbf{\Pi}} = diag(\overline{\boldsymbol{\pi}})$. Here, $\overline{\boldsymbol{\pi}}$ is a component-wise upper bound of the optimal value of the dual variable $\boldsymbol{\pi}(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y}$. Note that the matrix $\overline{\mathbf{\Pi}}$ is similar to M of the Big-M method in that it estimates an upper bound to the dual variables. We provide a method to estimate $\overline{\boldsymbol{\pi}}$ in Proposition 2.4.2. The dual problems of (P) and (P') are given by:

$$
\begin{aligned}
& g(\mathbf{x}, \mathbf{y}) = \\
\text{(D)} \quad & \min_{\boldsymbol{\pi}, \mathbf{q}} \ \mathbf{q}^\top \mathbf{d} + \boldsymbol{\pi}^\top \mathbf{v} + \boldsymbol{\pi}^\top \mathbf{W}(\mathbf{e} - \mathbf{x}) \\
& \text{s.t.} \ \ \boldsymbol{\pi}^\top + \mathbf{q}^\top \mathbf{D} \geq \mathbf{y}^\top \\
& \qquad \boldsymbol{\pi}, \mathbf{q} \geq \mathbf{0},
\end{aligned}
\qquad
\begin{aligned}
& \bar{g}(\mathbf{x}, \mathbf{y}) = \\
& \min_{\mathbf{r}, \mathbf{s}, \mathbf{t}} \ \mathbf{t}^\top \mathbf{d} + \mathbf{r}^\top \mathbf{W}\mathbf{e} + \mathbf{s}^\top \mathbf{v} \\
\text{(D')} \quad & \text{s.t.} \ \ \mathbf{s}^\top + \mathbf{t}^\top \mathbf{D} \geq \mathbf{y}^\top \\
& \qquad \mathbf{r}^\top + \mathbf{t}^\top \mathbf{D} \geq \mathbf{y}^\top - \mathbf{x}^\top \overline{\mathbf{\Pi}} \\
& \qquad \mathbf{r}, \mathbf{s}, \mathbf{t} \geq \mathbf{0}.
\end{aligned}
$$

Proposition 2.4.1. *Given a binary* $\mathbf{x}$*, if the set* $\mathcal{U}^{\overline{\mathbf{\Pi}}}(\mathbf{x})$ *is nonempty and* $\mathbf{v}, \mathbf{W} \geq 0$*, then for all* $\mathbf{y}$*:*

$$h(\mathbf{x}, \mathbf{y}) = \bar{h}(\mathbf{x}, \mathbf{y}).$$

Proof. Strong duality warrants the equalities $g(\mathbf{x}, \mathbf{y}) = h(\mathbf{x}, \mathbf{y})$ and $\bar{g}(\mathbf{x}, \mathbf{y}) = \bar{h}(\mathbf{x}, \mathbf{y})$. In the following, we also refer to the optimal objective values of the dual problems as $h(\mathbf{x}, \mathbf{y})$ and $\bar{h}(\mathbf{x}, \mathbf{y})$. Let $(\boldsymbol{\pi}, \mathbf{q})$ be an optimal solution to (D). Furthermore, let $(\mathbf{r} = \boldsymbol{\pi} - \mathbf{\Pi}\mathbf{x}, \mathbf{s} = \boldsymbol{\pi}, \mathbf{t} = \mathbf{q})$ with $\mathbf{\Pi} = diag(\boldsymbol{\pi})$ be a potential feasible solution to (D'). For these solutions, it follows that $\mathbf{s}^\top + \mathbf{t}^\top \mathbf{D} = \boldsymbol{\pi}^\top + \mathbf{q}^\top \mathbf{D} \geq \mathbf{y}^\top$, and
$\mathbf{r}^\top + \mathbf{t}^\top \mathbf{D} = \boldsymbol{\pi}^\top - \mathbf{x}^\top \mathbf{\Pi} + \mathbf{q}^\top \mathbf{D} \geq \mathbf{y}^\top - \mathbf{x}^\top \mathbf{\Pi} \geq \mathbf{y}^\top - \mathbf{x}^\top \overline{\mathbf{\Pi}}$. Since $\boldsymbol{\pi}, \mathbf{q} \geq \mathbf{0}$, and $\mathbf{x}$ is binary, we obtain $\mathbf{r}, \mathbf{s}, \mathbf{t} \geq \mathbf{0}$. This means $(\mathbf{r}, \mathbf{s}, \mathbf{t})$ is a feasible solution to problem (D'). This

yields

$$\begin{aligned}\bar{h}(\mathbf{x},\mathbf{y}) &\leq \mathbf{q}^\top\mathbf{d} + \boldsymbol{\pi}^\top\mathbf{v} + (\boldsymbol{\pi} - \boldsymbol{\Pi}\mathbf{x})^\top\mathbf{W}\mathbf{e} \\ &= h(\mathbf{x},\mathbf{y}).\end{aligned}$$

For the converse, let $(\mathbf{r},\mathbf{s},\mathbf{t})$ be an optimal solution to (D'). Consider $(\boldsymbol{\pi}=\mathbf{s},\mathbf{q}=\mathbf{t})$ to be a solution to (D). The feasibility of $(\mathbf{r},\mathbf{s},\mathbf{t})$ leads $\boldsymbol{\pi}^\top + \mathbf{q}^\top\mathbf{D} = \mathbf{s}^\top + \mathbf{t}^\top\mathbf{D} \geq \mathbf{y}^\top$, and $\boldsymbol{\pi}=\mathbf{s}\geq\mathbf{0}, \mathbf{q}=\mathbf{t}\geq\mathbf{0}$. Hence, $(\boldsymbol{\pi},\mathbf{q})$ is a feasible solution to (D), resulting in

$$\begin{aligned}h(\mathbf{x},\mathbf{y}) &\leq \mathbf{t}^\top\mathbf{d} + \mathbf{s}^\top\mathbf{v} + \mathbf{s}^\top\mathbf{W}(\mathbf{e}-\mathbf{x}) \\ &= \bar{h}(\mathbf{x},\mathbf{y}) + (\mathbf{s}-\mathbf{r})^\top\mathbf{W}\mathbf{e} - \mathbf{s}^\top\mathbf{W}\mathbf{x}.\end{aligned}$$

In order to prove $h(\mathbf{x},\mathbf{y}) \leq \bar{h}(\mathbf{x},\mathbf{y})$, it is required to prove $(\mathbf{s}-\mathbf{r})^\top\mathbf{W}\mathbf{e} - \mathbf{s}^\top\mathbf{W}\mathbf{x} \leq 0$. This can be expressed as $\sum_i w_i(s_i - r_i - s_i x_i) \leq 0$. For all i with $x_i = 1$, it holds that $w_i(s_i - r_i - s_i x_i) = -w_i r_i \leq 0$.

Consider now the set of all i with $x_i = 0$, denoted by X_0. Problem (D') can be rewritten as two nested minimization problems, where the outer problem is over $\mathbf{t}$ and r_j, s_j with $j \notin X_0$ and the inner problem over r_i, s_i with $i \in X_0$:

$$\bar{h}(\mathbf{x},\mathbf{y}) = \min_{\mathbf{t},r_j,s_j,j\notin X_0} \mathbf{t}^\top\mathbf{d} + \sum_{j\notin X_0} r_j w_j + \sum_{j\notin X_0} s_j v_j + l(\mathbf{t})$$

$$\left.\begin{aligned}\text{s.t. } & s_j + \mathbf{t}^\top\mathbf{D}_{\cdot,j} \geq y_j \\ & r_j + \mathbf{t}^\top\mathbf{D}_{\cdot,j} \geq y_j - \overline{\pi}_j \\ & r_j, s_j \geq \mathbf{0}\end{aligned}\right\} \forall j \notin X_0.$$

The inner minimization is captured by the function $l(\mathbf{t})$, which is given by

$$l(\mathbf{t}) = \min_{r_i, s_i,\, i \in X_0} \sum_{i \in X_0} r_i w_i + \sum_{i \in X_0} s_i v_i$$
$$\left.\begin{aligned} \text{s.t. } & s_i + \mathbf{t}^\top \mathbf{D}_{\bullet,i} \geq y_i \\ & r_i + \mathbf{t}^\top \mathbf{D}_{\bullet,i} \geq y_i \\ & r_i, s_i \geq \mathbf{0} \end{aligned}\right\} \forall\, i \in X_0.$$

Note that in this inner minimization problem, the same constraints act on s_i and r_i. Since w_i and v_i are nonnegative, there exist optimal solutions s_i and r_i that are equal and set to their lower bounds $s_i = r_i = \max\{y_i - \mathbf{t}^\top \mathbf{D}_{\bullet,i}, 0\}$. Therefore, $\sum_{i \in X_0} s_i w_i - r_i w_i = 0$, which means $h(\mathbf{x}, \mathbf{y}) \leq \bar{h}(\mathbf{x}, \mathbf{y})$. □

Using Proposition 2.4.1 and problem (D'), the constraint (LC) can be reformulated as

$$\mathbf{t}^\top \mathbf{d} + \mathbf{r}^\top \mathbf{W}\mathbf{e} + \mathbf{s}^\top \mathbf{v} \leq b$$
$$\mathbf{s}^\top + \mathbf{t}^\top \mathbf{D} \geq \mathbf{y}^\top$$
$$\mathbf{r}^\top + \mathbf{t}^\top \mathbf{D} \geq \mathbf{y}^\top - \mathbf{x}^\top \overline{\mathbf{\Pi}}$$
$$\mathbf{r}, \mathbf{s}, \mathbf{t} \geq \mathbf{0}.$$

Note that this reformulation does not contain any bilinear terms and includes fewer constraints than the standard Big-M formulations. Additionally, Proposition 2.4.1 allows us to replace $h(\mathbf{x}, \mathbf{y})$ with $\bar{h}(\mathbf{x}, \mathbf{y})$. This is important because $\bar{h}(\mathbf{x}, \mathbf{y})$ is convex in $(\mathbf{x}, \mathbf{y})$.

Therefore, cut generation algorithms can be used to solve this problem which is not possible for the original problem with the constraint (LC). In the following, we discuss the matrix $\overline{\mathbf{\Pi}}$.

Estimation of $\overline{\mathbf{\Pi}}$ The following proposition sheds light on how to estimate $\overline{\mathbf{\Pi}}$.

Proposition 2.4.2. *If* $\mathbf{D}$ *and* $\mathbf{y}$ *are element-wise nonnegative, then* $\pi_i(\mathbf{x}, \mathbf{y}) \leq y_i$ $\forall (\mathbf{x}, \mathbf{y})$ *for constraint* (LC) *under the uncertainty set* $\mathcal{U}^{\overline{\mathbf{\Pi}}}$.

Proof. Consider the following problem for some index i

$$
\begin{aligned}
F(\theta) = \max_{\boldsymbol{\xi}} \ & \mathbf{y}^\top \boldsymbol{\xi} \\
\text{s.t.} \ & \mathbf{D}\boldsymbol{\xi} \leq \mathbf{d} && : \mathbf{q} \\
& \boldsymbol{\xi} \leq \mathbf{v} + \mathbf{W}(\mathbf{e} - \mathbf{x}) + \theta \mathbf{e}_i && : \boldsymbol{\pi} \\
& \boldsymbol{\xi} \geq \mathbf{0}.
\end{aligned} \tag{2.1}
$$

Let $\boldsymbol{\xi}_0$ be the optimal solution at $\theta = 0$ and the corresponding optimal dual variables are $\mathbf{q}_0$ and $\boldsymbol{\pi}_0$. Let the optimal basis of the above problem be given by some matrix $\mathbf{B}$. Since $\boldsymbol{\xi}_0$ is the optimal solution, the vector of basic variables is given by $\boldsymbol{\xi}_0^B = \mathbf{B}^{-1}\mathbf{b}$, where $\mathbf{b}$ denotes the RHS vector of problem (2.1), i.e., $\mathbf{b} = [\mathbf{d}^\top, \mathbf{v}^\top + (\mathbf{e} - \mathbf{x})^\top \mathbf{W}]^\top$. Assume that the solution is non-degenerate. This means $\mathbf{B}^{-1}\mathbf{b} > 0$. Then for a small enough change in $\mathbf{b}$, the optimal basis does not change. If it is degenerate, then $\mathbf{b}$ can be perturbed by a small ϵ to obtain a non-degenerate solution, which only marginally changes the objective [see, e.g., 19].

When $\theta > 0$ is small enough, the basis matrix does not change. This means that both solutions (corresponding to $\theta = 0$ and $\theta > 0$) have the same dual variables because the

dual variables do not depend on the RHS vector. This means

$$\begin{aligned} F(\theta) - F(0) = \quad & \boldsymbol{\pi}_0^\top \mathbf{v} + \boldsymbol{\pi}_0^\top \mathbf{W}(\mathbf{e} - \mathbf{x}) + \theta \boldsymbol{\pi}_0^\top \mathbf{e}_i + \mathbf{q}_0^\top \mathbf{d} - \boldsymbol{\pi}_0^\top \mathbf{v} - \boldsymbol{\pi}_0^\top \mathbf{W}(\mathbf{e} - \mathbf{x}) - \mathbf{q}_0^\top \mathbf{d} \\ = \quad & \theta \boldsymbol{\pi}_0^\top \mathbf{e}_i, \end{aligned}$$

which represents the change in the objective value. Let $\boldsymbol{\xi}_0$ be the optimal solution of the problem with $\theta = 0$ and $\boldsymbol{\xi}_\theta$ be the optimal solution of problem with $\theta > 0$. Then the change in the objective value is

$$\theta \boldsymbol{\pi}_0^\top \mathbf{e}_i = \mathbf{y}^\top \boldsymbol{\xi}_\theta - \mathbf{y}^\top \boldsymbol{\xi}_0.$$

Using Lemma 2.4.1, we can state that

$$\begin{aligned} \theta \boldsymbol{\pi}_0^\top \mathbf{e}_i &= \mathbf{y}^\top \boldsymbol{\xi}_\theta - \mathbf{y}^\top \boldsymbol{\xi}_0 \\ &\leq \mathbf{y}^\top \boldsymbol{\xi}_0 + \theta \mathbf{y}^\top \mathbf{e}_i - \mathbf{y}^\top \boldsymbol{\xi}_0 \\ &= \theta \mathbf{y}^\top \mathbf{e}_i. \end{aligned}$$

This implies that $\pi_{0,i} \leq y_i \ \ \forall i$. □

Corollary 2.4.1. *Proposition 2.4.2 allows the estimation of* $\overline{\Pi}$ *by*

$$\begin{aligned} \overline{\pi}_i = \max_{\mathbf{y}} \ & \mathbf{y}^\top \mathbf{e}_i \\ s.t. \ & (\mathbf{x}, \mathbf{y}) \in Y \\ & x_i \in \{0, 1\}, \end{aligned} \tag{2.2}$$

where set Y *denotes the remaining constraints of the original full problem.*

Lemma 2.4.1. *If the matrix* $\mathbf{D}$ *is element-wise greater than* 0*, then* $\boldsymbol{\xi}_\theta \leq \boldsymbol{\xi}_0 + \theta \mathbf{e}_i$.

Proof. Suppose this is not true, i.e., there exists at least one index k such that $\xi_{\theta,k} > \xi_{0,k} + \theta e_{i,k}$. In addition, it holds that for $\theta \geq 0$, $\mathbf{y}^\top \boldsymbol{\xi}_\theta > \mathbf{y}^\top \boldsymbol{\xi}_0$.
If $k \neq i$, then $\boldsymbol{\xi}_\theta \leq \mathbf{v} + \mathbf{W}(\mathbf{e} - \mathbf{x})$, which suggests $\boldsymbol{\xi}_\theta$ to be feasible for the problem with $\theta = 0$. This would contradict the optimality of $\boldsymbol{\xi}_0$.
If $k = i$, then $\xi_{\theta,i} > \xi_{0,i} + \theta$. However, this results in $\boldsymbol{\xi}_0 < \boldsymbol{\xi}_\theta - \theta \mathbf{e}_i \leq \mathbf{v} + \mathbf{W}(\mathbf{e} - \mathbf{x})$. Since $D(\boldsymbol{\xi}_\theta - \theta \mathbf{e}_i) = \mathbf{D}\boldsymbol{\xi}_\theta - \theta \mathbf{D}\mathbf{e}_i \leq \mathbf{d} - \theta \mathbf{D}\mathbf{e}_i \leq \mathbf{d}$, $\boldsymbol{\xi}_\theta - \theta \mathbf{e}_i$ is a feasible solution to the problem with $\theta = 0$. However, this indicates that $\mathbf{y}^\top(\boldsymbol{\xi}_\theta - \theta \mathbf{e}_i) > \mathbf{y}^\top \boldsymbol{\xi}_0$ which also contradicts the optimality of $\boldsymbol{\xi}_0$. Therefore, we can conclude that $\boldsymbol{\xi}_\theta \leq \boldsymbol{\xi}_0 + \theta \mathbf{e}_i$. □

This proposition allows us to estimate $\overline{\pi}_i$ by setting it equal to the maximum value that y_i can take in the overall problem. In some cases, such as shortest path or facility location problems, this is straightforwardly estimated from the underlying model. With this, all components of the decision-dependent problem with the polyhedral uncertainty set $\mathcal{U}^{\overline{\mathbf{\Pi}}}$ can be computed efficiently for practical size problems. We now extend Proposition 2.4.1 to more general uncertainty sets.

2.4.2. Extension to conic sets

Given a convex cone $\mathcal{K}$, the decision-dependent uncertainty set $\mathcal{U}^{\overline{\mathbf{\Pi}}}(\mathbf{x})$ can be extended to

$$\mathcal{U}^{\mathcal{K}}(\mathbf{x}) = \{\boldsymbol{\xi} \mid \mathbf{d} - \mathbf{D}\boldsymbol{\xi} \in \mathcal{K},\ \boldsymbol{\xi} \leq \mathbf{v} + \mathbf{W}(\mathbf{e} - \mathbf{x}), \boldsymbol{\xi} \geq \mathbf{0}\}.$$

Here $\mathbf{d}$ and $\mathbf{D}$ are coefficients and $\mathbf{v}$ and $\mathbf{W} = diag(\mathbf{w})$ denote upper bounds to the uncertain component $\boldsymbol{\xi}$. The objective is to reformulate the constraint $\mathbf{y}^\top \boldsymbol{\xi} \leq b$, $\forall \boldsymbol{\xi} \in$

$\mathcal{U}^{\mathcal{K}}(\mathbf{x})$. In order to satisfy this constraint for all $\boldsymbol{\xi} \in \mathcal{U}^{\mathcal{K}}(\mathbf{x})$, its LHS can be expressed with the following two problems:

$$
\begin{aligned}
& h(\mathbf{x}, \mathbf{y}) = \\
& \max_{\boldsymbol{\xi}} \ \mathbf{y}^{\top} \boldsymbol{\xi} \\
\text{(KP)} \quad & \text{s.t. } \mathbf{d} - \mathbf{D}\boldsymbol{\xi} \in \mathcal{K} \\
& \boldsymbol{\xi} \leq \mathbf{v} + \mathbf{W}(\mathbf{e} - \mathbf{x}) \ : \ \boldsymbol{\pi}(\mathbf{x}, \mathbf{y}) \\
& \boldsymbol{\xi} \geq \mathbf{0},
\end{aligned}
$$

$$
\begin{aligned}
& \bar{h}(\mathbf{x}, \mathbf{y}) = \\
& \max_{\boldsymbol{\xi}, \boldsymbol{\zeta}} \ (\mathbf{y} - \overline{\boldsymbol{\Pi}}\mathbf{x})^{\top} \boldsymbol{\xi} + \mathbf{y}^{\top} \boldsymbol{\zeta} \\
\text{(KP')} \quad & \text{s.t. } \mathbf{d} - \mathbf{D}\boldsymbol{\xi} \in \mathcal{K} \\
& \boldsymbol{\xi} \leq \mathbf{We} \\
& \boldsymbol{\zeta} \leq \mathbf{v} \\
& \boldsymbol{\xi}, \boldsymbol{\zeta} \geq \mathbf{0}.
\end{aligned}
$$

Here, $\boldsymbol{\pi}(\mathbf{x}, \mathbf{y})$ denotes the dual variable for the corresponding constraint. Let $\overline{\boldsymbol{\Pi}}$ be an element-wise upper bound on the dual variables $\boldsymbol{\pi}(\mathbf{x}, \mathbf{y})$. The following proposition shows that the problems (KP) and (KP') have the same optimal objective value.

Proposition 2.4.3. *If* $\forall \mathbf{x} \in \{0,1\}^n$ *there exists a point* $\bar{\boldsymbol{\xi}}$ *such that* $\mathbf{d} - \mathbf{D}\bar{\boldsymbol{\xi}}$ *lies in the relative interior of* $\mathcal{K}$ *and* $\mathbf{0} \leq \bar{\boldsymbol{\xi}} \leq \mathbf{v} + \mathbf{W}(\mathbf{e} - \mathbf{x})$ *with* $\mathbf{v}, \mathbf{W} \geq 0$*, then for all* $\mathbf{x}, \mathbf{y}$*:*

$$h(\mathbf{x}, \mathbf{y}) = \bar{h}(\mathbf{x}, \mathbf{y}).$$

The proof of this proposition proceeds similar to that of Proposition 2.4.1. Strong duality holds due to the assumptions on the uncertainty set [15, Proposition 5.3.1]. This proposition allows us to utilize the convex counterpart of the function h in what follows. For a complete proof, refer to [82].

Using Proposition 2.4.3 and the dual problem of (KP'), the constraint (LC) can be reformulated as

$$\mathbf{t}^\top \mathbf{d} + \mathbf{r}^\top \mathbf{W}\mathbf{e} + \mathbf{s}^\top \mathbf{v} \leq b$$

$$\mathbf{s}^\top + \mathbf{t}^\top \mathbf{D} \geq \mathbf{y}^\top$$

$$\mathbf{r}^\top + \mathbf{t}^\top \mathbf{D} \geq \mathbf{y}^\top - \mathbf{x}^\top \overline{\mathbf{\Pi}}$$

$$\mathbf{t} \in \mathcal{K}^*,\ \mathbf{r}, \mathbf{s} \geq \mathbf{0},$$

with the dual cone $\mathcal{K}^*$. Note that this reformulation has only linear terms and, as we will see in Section 2.5, fewer constraints than the standard Big-M formulation, hence it is more suitable for larger sized problems. The proof of this formulation proceeds parallel to that of Proposition 2.4.1.

In summary, these results allow the modeling of uncertainty sets with reducible upper bounds. Such bounds motivate the notion of *proactive uncertainty control*. It mitigates conservatism and better actualizes the tradeoff between cost of control and disadvantage of uncertainty, both of which are instrumental parts of many real-world applications. Until now, we discussed the special polyhedral set $\mathcal{U}^{\overline{\mathbf{\Pi}}}$. The following section provides a reformulation of problem (RO-DDU) under general polyhedral uncertainty sets.

2.5. Extensions to General Polyhedral Sets

The previous section leveraged the specific structure of the uncertainty set to obtain smaller reformulations. The Big-M reformulation, however, has the advantage of not requiring any special set structure. For completeness and a comparison of formulation sizes, the following proposition reformulates problem (RO-DDU) for the general polyhedral set $\mathcal{U}^P(\mathbf{x})$.

Proposition 2.5.1. *If the uncertainty set* $\mathcal{U}_i(\mathbf{x})$ *is a polyhedron as in* $\mathcal{U}^P(\mathbf{x})$ *with* $\boldsymbol{D}_i \in \mathbb{R}^{m_i \times p}$, $\mathbf{d}_i \in \mathbb{R}^{m_i}$, *and* $\boldsymbol{\Delta}_i \in \mathbb{R}^{m_i \times n}$ *and if* $\mathbf{x}$ *is binary, then the robust counterpart of problem* (RO-DDU) *is*

$$
\begin{aligned}
&\min_{\mathbf{x},\mathbf{y},\mathbf{w},\boldsymbol{\pi}} \quad \mathbf{c}^\top \mathbf{x} + \mathbf{f}^\top \mathbf{y} \\
&\left.\begin{aligned} s.t.\ \ & \mathbf{a}_i^\top \mathbf{x} + \boldsymbol{\pi}_i^\top \mathbf{d}_i + \sum_{j=1}^{m_i} \sum_{k=1}^{n} \Delta_{ijk} w_{ijk} \leq b_i \\ & \boldsymbol{\pi}_i^\top \mathbf{D}_i = \mathbf{y}^\top \end{aligned}\right\} \quad \forall i \\
&\left.\begin{aligned} & w_{ijk} \leq M x_k, \ \ w_{ijk} \leq \pi_{ij} \\ & w_{ijk} \geq \pi_{ij} - M(1 - x_k) \\ & \pi_{ij} \geq 0, \ w_{ijk} \geq 0 \end{aligned}\right\} \quad \forall i, j, k \\
&\quad \mathbf{x} \in \{0,1\}^n,
\end{aligned}
$$

where M *is a sufficiently large number.*

Proof. We consider two cases, namely: *Case 1:* There exists a feasible solution $(\mathbf{x}, \mathbf{y})$ to (RO-DDU). Therefore, $\mathbf{x}$ and $\mathbf{y}$ must satisfy all constraints $\mathbf{a}_i^\top \mathbf{x} + \boldsymbol{\xi}_i^\top \mathbf{y} \leq b_i \ \forall \boldsymbol{\xi}_i \in \mathcal{U}_i(\mathbf{x})$ for all i. This is equivalent to

$$\mathbf{a}_i^\top \mathbf{x} + \max_{\boldsymbol{\xi}_i \in \mathcal{U}_i(\mathbf{x})} \boldsymbol{\xi}_i^\top \mathbf{y} \leq b_i \quad \forall i. \tag{2.3}$$

If this problem is feasible and has a finite optimal solution, then by strong duality, the corresponding dual problem has the same objective value. Problem (2.3) can now be

expressed as

$$\left.\begin{aligned} &\mathbf{a}_i^\top \mathbf{x} + \boldsymbol{\pi}_i^\top (\mathbf{d}_i + \boldsymbol{\Delta}_i \mathbf{x}) \leq b_i \\ &\boldsymbol{\pi}_i^\top \mathbf{D}_i = \mathbf{y}^\top \\ &\boldsymbol{\pi}_i \geq \mathbf{0} \end{aligned}\right\} \forall i, \tag{2.4}$$

where $\boldsymbol{\pi}_i \in \mathbb{R}^{m_i}$ is the dual variable for constraints corresponding to the uncertainty set $\mathcal{U}_i(\mathbf{x})$. Here m_i refers to the number of constraints in the set $\mathcal{U}_i(\mathbf{x})$. Since the primal problem is feasible and finitely valued, there exists a $\boldsymbol{\pi}_i$, for which the constraints (2.4) are satisfied. Therefore, the original problem (RO-DDU) can be written as

$$\begin{aligned} &\min_{\boldsymbol{\pi}_i, \mathbf{x}, \mathbf{y}} \ \mathbf{c}^\top \mathbf{x} + \mathbf{f}^\top \mathbf{y} \\ &\left.\begin{aligned} &\text{s.t. } \mathbf{a}_i^\top \mathbf{x} + \boldsymbol{\pi}_i^\top \mathbf{d}_i + \boldsymbol{\pi}_i^\top \boldsymbol{\Delta}_i \mathbf{x} \leq b_i \\ &\boldsymbol{\pi}_i^\top \mathbf{D}_i = \mathbf{y}^\top \\ &\boldsymbol{\pi}_i \geq \mathbf{0} \end{aligned}\right\} \forall i. \end{aligned}$$

Note the bilinear term in the first constraint. By expanding the variable space, the ith constraint can be rewritten as

$$\mathbf{a}_i^\top \mathbf{x} + \sum_{j=1}^{m_i} \pi_{ij} d_{ij} + \sum_{j=1}^{m_i} \sum_{k=1}^{n} \Delta_{ijk} w_{ijk} \leq b_i, \ \text{ with } w_{ijk} = \pi_{ij} x_k.$$

In the bilinear term, $w_{ijk} = \pi_{ij} x_k$, x_k is binary, allowing to rewrite the term as

$$w_{ijk} \leq \pi_{ij}, \ \ 0 \leq w_{ijk} \leq M x_k, \ \ w_{ijk} \geq \pi_{ij} - M(1 - x_k),$$

where M is a sufficiently large constant. Consequently, the problem (RO-DDU) can be reformulated as

$$
\begin{aligned}
&\min_{\mathbf{x},\mathbf{y}} \quad \mathbf{c}^\top \mathbf{x} + \mathbf{f}^\top \mathbf{y} \\
&\left.\begin{aligned}
&\text{s.t.} \quad \mathbf{a}_i^\top \mathbf{x} + \boldsymbol{\pi}_i^\top \mathbf{d}_i + \sum_{j=1}^{m_i} \sum_{k=1}^{n} \Delta_{ijk} w_{ijk} \leq b_i \\
&\qquad \boldsymbol{\pi}_i^\top \mathbf{D}_i = \mathbf{y}^\top
\end{aligned}\right\} \quad \forall i \\
&\left.\begin{aligned}
&w_{ijk} \leq M x_k, \quad w_{ijk} \leq \pi_{ij} \\
&w_{ijk} \geq \pi_{ij} - M(1 - x_k) \\
&\boldsymbol{\pi}_i \geq \mathbf{0}, \; w_{ijk} \geq 0
\end{aligned}\right\} \quad \forall i, j, k \\
&\qquad \mathbf{x} \in \{0,1\}^n.
\end{aligned}
\tag{2.5}
$$

Case 2: Problem (RO-DDU) is infeasible. Then the reformulation in (2.5) is infeasible. To show this, consider the original problem (RO-DDU).

Suppose this problem is infeasible under the assumptions of Proposition 2.5.1. This means that $\forall \mathbf{x} : \exists \boldsymbol{\xi} \in \mathcal{U}(\mathbf{x})$ such that $\mathbf{a}_i^\top \mathbf{x} + \boldsymbol{\xi}_i^\top \mathbf{y} > b_i$. Consequently, the constraint $\mathbf{a}_i^\top \mathbf{x} + \max_{\boldsymbol{\xi}_i \in \mathcal{U}_i(\mathbf{x})} \boldsymbol{\xi}_i^\top \mathbf{y} > b_i$ holds for at least one i. Using the dual of the inner problem, the constraints can be written $\forall \boldsymbol{\pi}_i$ as

$$
\begin{aligned}
&\mathbf{a}_i^\top \mathbf{x} + \boldsymbol{\pi}_i^\top (\mathbf{d}_i + \boldsymbol{\Delta}_i \mathbf{x}) > b_i \\
&\boldsymbol{\pi}_i^\top \mathbf{D}_i = \mathbf{y}^\top \\
&\boldsymbol{\pi}_i \geq \mathbf{0}.
\end{aligned}
\tag{2.6}
$$

Now, assume that the reformulation in (2.5) is feasible. Given its constraints, there exists a binary vector $\mathbf{x}$ and a vector $\mathbf{w}$ such that $w_{ijk} = \pi_{ij}x_k$. However, this implies a variable $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{ik}, \ldots, \pi_{im_i})$ that satisfies $\boldsymbol{\pi}_i^\top \mathbf{D}_i = \mathbf{y}^\top$, $\boldsymbol{\pi}_i \geq \mathbf{0}$ and

$$\mathbf{a}_i^\top \mathbf{x} + \sum_{j=1}^{m_i} \pi_{ij} d_{ij} + \sum_{j=1}^{m_i} \sum_{k=1}^{n} \Delta_{ijk} \pi_{ij} x_k \leq b_i.$$

This contradicts the earlier assertion in (2.6) that there exist no such $\boldsymbol{\pi}_i$. □

This proposition allows us to reformulate the original decision-dependent RO problem as a mixed integer linear program which can be solved for many realistic size problems using off-the-shelf algorithms. Such mixed integer reformulations can also be provided for general convex uncertainty sets [14], which includes conic and budgeted structures. Their proofs (not shown) proceed parallel to that of Proposition 2.5.1.

Note that problem (RO-DDU) has n binary and p continuous variables, along with m constraints. The i$^{\text{th}}$ uncertain $\boldsymbol{\xi}_i$ lies in an uncertainty set with m_i constraints. Table 2.1 presents the size of the reformulation under two settings: (i) $\mathbf{x}$ is binary as in Proposition 2.5.1 and (ii) x_i can take s possible values. For the sake of clarity, we assume that $m_i = K \ \forall i$, where K is some constant. Table 2.1 shows that for (ii), the size of the reformulation increases rapidly with growing s. In certain cases, it is possible to improve the Big-M reformulation by imposing mild assumptions, as we will discuss next.

Nature of $\mathbf{x}$	Binary var.	Continuous var.	Affine constr.	Sign constr.
Binary	n	$p + mK + nK$	$m + mp + 3nK$	$mK(n+1)$
Finite valued	$(n+1)s$	$p + mK$ $+nmK(s+1)$	$m + mp + 2n$ $+nmK(3s+1)$	$mK(ns+1)$

Table 2.1. Size of Big-M formulation of (RO-DDU) for $\mathcal{U}_i(\mathbf{x})$ with respect to (i) $\mathbf{x} \in \{0,1\}^n$ and (ii) $\mathbf{x} \in \mathbb{R}^n$ with x_i taking s possible values: $\dim(\mathbf{y}) = p$, K constraints in $\mathcal{U}_i(\mathbf{x})$, and m constraints in the complete problem.

2.5.1. Modified Big-M Reformulation

Consider the uncertainty set $\mathcal{U}^{\mathrm{P}}(\mathbf{x})$ to be expressed as

$$\mathcal{U}^{\mathrm{P}}(\mathbf{x}) = \left\{ \boldsymbol{\xi} \mid \mathbf{D}_{i\cdot}^{\top}\boldsymbol{\xi} \leq d_i + \sum_{j=1}^{n} \Delta_{ij} x_j, \ \ \forall i = 1, \ldots, m \right\}.$$

To overcome the poor numerical performance of standard Big-M reformulation due to its weak relaxations, we impose the mild assumption that all elements of the coefficient matrix $\boldsymbol{\Delta}$ are non-negative. Proposition 2.5.2 reformulates constraint (LC) for $\mathcal{U}^{\mathrm{P}}(\mathbf{x})$ under this assumption.

Proposition 2.5.2. *If $\Delta_{ij} \geq 0 \ \ \forall i, j$, then the constraint* (LC) *with the uncertainty set $\mathcal{U}^{P}(\mathbf{x})$ and a large constant M can be reformulated as*

$$\left.\begin{array}{ll} \displaystyle\sum_{i=1}^{m} \pi_i d_i + \sum_{i=1}^{m}\sum_{j=1}^{n} t_{ij} \leq b & \\ \displaystyle\sum_{i=1}^{m} \pi_i D_{ij} = y_j, & \forall j \\ t_{ij} \geq \pi_i \Delta_{ij} - M(1 - x_j) & \\ \pi_i \geq 0, \ \ t_{ij} \geq 0 & \end{array}\right\} \ \forall i, j.$$

Proof. The LHS maximization problem for the constraint (LC) can be written as

$$\begin{aligned} \max_{\boldsymbol{\xi}} \ & \mathbf{y}^{\top}\boldsymbol{\xi} \\ \text{s.t.} \ & \mathbf{D}_{i\cdot}^{\top}\boldsymbol{\xi} \leq d_i + \sum_{j=1}^{n} \Delta_{ij} x_j \ \ \forall i. \end{aligned}$$

Using the dual of this problem, the original constraint $\mathbf{y}^\top \boldsymbol{\xi} \leq b \quad \forall \boldsymbol{\xi} \in \mathcal{U}^{\mathrm{P}}(\mathbf{x})$ can be written as

$$\begin{aligned} & \sum_{i=1}^{m} \pi_i (d_i + \sum_{j=1}^{n} \Delta_{ij} x_j) \leq b \\ & \sum_{i=1}^{m} \pi_i D_{ij} = y_j \quad \forall j \\ & \boldsymbol{\pi} \geq \mathbf{0}. \end{aligned} \tag{2.7}$$

The constraints in (2.7) can be rewritten by expanding the variable space as

$$\begin{aligned} & \sum_{i=1}^{m} \pi_i d_i + \sum_{i=1}^{m} \sum_{j=1}^{n} t_{ij} \leq b \\ & \pi_i \Delta_{ij} x_j \leq t_{ij} && \forall i, j \\ & \sum_{i=1}^{m} \pi_i D_{ij} = y_j && \forall j \\ & \boldsymbol{\pi} \geq \mathbf{0}. \end{aligned} \tag{2.8}$$

If there is a variable $\boldsymbol{\pi}$ feasible for the set of equations given by (2.7), then we can find a feasible variable for (2.8) by $t_{ij} = \pi_i \Delta_{ij} x_j$. On the other hand, if there exists a feasible solution to (2.8), then it is also feasible for (2.7). If $x_j = 0$, then $t_{ij} \geq 0$ and if $x_j = 1$, then $t_{ij} \geq \pi_i \Delta_{ij}$. This can be expressed as the following set of constraints

$$0 \leq t_j \geq \pi_i \Delta_{ij} - M(1 - x_j).$$

which completes the proof. □

The Proposition 2.5.2 leverages the fact that the variable t_{ij} remains at its lower bound, making the upper bounding constraints from the Big-M linearization redundant. However, if t_{ij} can be negative, the two lower bounding constraints are not sufficient. In some cases, it is possible to reformulate the problem even if the RHS coefficients

Formulations	Problem	Variables / Constraints
$\overline{\mathbf{\Pi}}$	$\mathbf{t}^\top\mathbf{d} + \mathbf{r}^\top\mathbf{W}\mathbf{e} + \mathbf{s}^\top\mathbf{v} \le b$ $\mathbf{s}^\top + \mathbf{t}^\top\mathbf{D} \ge \mathbf{y}^\top$ $\mathbf{r}^\top + \mathbf{t}^\top\mathbf{D} \ge \mathbf{y}^\top - \mathbf{x}^\top\overline{\mathbf{\Pi}}$ $\mathbf{r}, \mathbf{s}, \mathbf{t} \ge \mathbf{0}$.	C: $m + 2n$ A: $1 + 2n$ S: $m + 2n$
Big-M	$\mathbf{t}^\top\mathbf{d} + \mathbf{s}^\top\mathbf{v} + \boldsymbol{s}^\top\mathbf{W}\mathbf{e} - \sum_i r_i \le b$ $\mathbf{s}^\top + \mathbf{t}^\top\mathbf{D} \ge \mathbf{y}^\top$ $w_i s_i - M(1 - x_i) \le r_i \le M x_i$ $r_i \le w_i s_i$ $\mathbf{r}, \mathbf{s}, \mathbf{t} \ge \mathbf{0}$.	C: $m + 2n$ A: $1 + 4n$ S: $m + 2n$
Modified Big-M	$\mathbf{t}^\top\mathbf{d} + \mathbf{s}^\top\mathbf{v} + \mathbf{r}^\top\mathbf{e} \le b$ $\mathbf{s}^\top + \mathbf{t}^\top\mathbf{D} \ge \mathbf{y}^\top$ $r_i \ge w_i s_i - M x_i$ $\mathbf{r}, \mathbf{s}, \mathbf{t} \ge \mathbf{0}$.	C: $m + 2n$ A: $1 + 2n$ S: $m + 2n$

Table 2.2. Comparison of (LC) reformulations for the set $\mathcal{U}^{\overline{\mathbf{\Pi}}}(\mathbf{x})$ (C: Continuous, A: Affine, S: Sign).

are negative. Consider the shortest path example presented in the introduction, which has constraints of the form $\xi_e \le 1 - \gamma_e x_e$. Here, the coefficient $\Delta_e = -\gamma_e$ is negative. However, we can rewrite the constraint as $\xi_e \le (1 - \gamma_e) + \gamma_e(1 - x_e)$ and apply the Big-M linearization on the variable $(1 - x_e)$ instead of on x_e. This substitution allows the use of the modified Big-M reformulation in more general settings. We report the numerical performance of this approach in comparison with the earlier reformulations in Section 2.6. For a comparison, we reformulate the constraint (LC) over the uncertainty set $\mathcal{U}^{\overline{\mathbf{\Pi}}}(\mathbf{x})$

using all three presented techniques, namely (i) $\overline{\mathbf{\Pi}}$, (ii) Big-M, and (iii) Modified Big-M. Table 2.2 presents this comparison along with the corresponding problem sizes. The sign constraints correspond to $(\bullet \geq 0)$, which are presented separately since they can be solved more efficiently. It displays that the primary difference between the Big-M and the other two reformulations is the larger number of affine (linear) constraints. To gain intuition and provide computational comparison between the different formulations, we extend the introductory example of Section 2.1 to a more detailed numerical experiment.

2.6. Numerical Experiments

Shortest path problems on networks constitute a general class of models, describing the most efficient connection between a source and target. Deterministic shortest routing problems can be solved with polynomial time algorithms [37]. However, this does not hold for uncertain arc lengths. Past research on robust shortest path problems focused on scenario-based [103], cardinality [17], and interval uncertainty [7, 106]. Despite a large body of literature, to the best of our knowledge, there is no work in the context of uncertainties that depend on decisions. To this end, our goals are:

(1) Comparing the numerical performance of different robust formulations,

(2) Measuring the benefit of proactive reduction as a function of size, budget, or cost of reduction,

(3) Measuring the number of arcs in the shortest path as a function of size, budget, or cost,

(4) Evaluating the price of robustness and the benefit of interacting with uncertainties, and

(5) Comparing the average and worst-case cost of decision dependence for RO and SO.

Here, we aim to model challenges that arise, e.g., in scenario planning of natural disasters. When sections of a transportation network are damaged, the actual travel times along arcs become uncertain. To plan for such a scenario, a decision-dependent RO solution can determine the arcs which should be strengthened (by reducing uncertainty) in order to improve the performance in an actual disaster. This strengthening incurs a fee. This means that it is possible to mitigate the impact of a disaster by managing the damage of a few particular arcs. Similarly, for transportation problems (e.g., air, ground), travel time can be improved by acquiring additional traffic or weather information on segments of the network.

To illustrate this setting, we discuss a problem on a graph $G = (\mathcal{V}, \mathcal{A}, d(\bullet))$ for the set of nodes $\mathcal{V}$, arcs $\mathcal{A}$, and the distance function $d(\bullet)$. The objective is to find the shortest path from the source to the target node ($s \to t$) when the actual realized distances from node i to j are uncertain and a function $d_{ij}(\boldsymbol{\xi}) = \left(1 + \frac{1}{2}\,\xi_{ij}\right)\bar{d}_{ij}$ of $\boldsymbol{\xi}$. The variable x_{ij} decides whether to reduce the maximum uncertainty in d_{ij}. This inquiry comes at a cost c_{ij}, which can be motivated as an investment in road improvement and is imposed on travelers via taxes or tolls. The parameter $\boldsymbol{\xi}$ resides in a cardinality constrained uncertainty set with reducible upper bounds. The complete problem is given by

$$\text{(SP)} \qquad \min_{\mathbf{x},\mathbf{y}} \max_{\boldsymbol{\xi}\in\mathcal{U}^{SP}(\mathbf{x})} \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} + \sum_{(i,j)\in\mathcal{A}} d_{ij}(\boldsymbol{\xi})y_{ij}$$
$$\text{s.t. } \mathbf{x} \in X \subseteq \{0,1\}^{|\mathcal{A}|},\ \mathbf{y} \in Y,$$

where y_{ij} decides whether the arc (i,j) lies in the shortest path. X denotes any constraints on $\mathbf{x}$ and Y the set of routing constraints. The uncertainty set is given by

$$\mathcal{U}^{SP}(\mathbf{x}) = \left\{ \boldsymbol{\xi} \mid \sum_{(i,j)\in\mathcal{A}} \xi_{ij} \le \Gamma,\ \xi_{ij} \le\ 1-\gamma_{ij}x_{ij},\ \xi_{ij} \ge 0\ \ \forall (i,j)\in\mathcal{A} \right\}.$$

We solve problem (SP) using the three different formulations: (i) $\overline{\mathbf{\Pi}}-$formulation from Proposition 2.4.1, (ii) standard Big-M formulation, and (iii) Modified Big-M formulation from Proposition 2.5.2. In Table 2.3, $X \times Y$ denote the collection of both the shortest path

Form.	Problem	Variables / Constraints
$\overline{\mathbf{\Pi}}$	$\min\limits_{\substack{\mathbf{x},\mathbf{y}\\ \mathbf{q},\mathbf{r},p}} f(\mathbf{x},\mathbf{y})+p\Gamma+\sum\limits_{(i,j)\in\mathcal{A}} q_{ij}(1-\gamma_{ij})+\sum\limits_{(i,j)\in\mathcal{A}} r_{ij}\gamma_{ij}$ s.t. $p+q_{ij} \ge \dfrac{y_{ij}d_{ij}-\overline{\pi}_{ij}d_{ij}x_{ij}}{2}$ $p+r_{ij} \ge \dfrac{y_{ij}d_{ij}}{2}$ $p, q_{ij}, r_{ij} \ge 0,\ \mathbf{x},\mathbf{y} \in X\times Y.$	B: $2\lvert\mathcal{A}\rvert$ C: $2\lvert\mathcal{A}\rvert+1$ A: $\lvert\mathcal{V}\rvert+2\lvert\mathcal{A}\rvert$ S: $2\lvert\mathcal{A}\rvert+1$
Big-M	$\min\limits_{\substack{\mathbf{x},\mathbf{y}\\ \mathbf{q},\mathbf{r},p}} f(\mathbf{x},\mathbf{y})+p\Gamma+\sum\limits_{(i,j)\in\mathcal{A}} q_{ij}-\sum\limits_{(i,j)\in\mathcal{A}} \gamma_{ij}r_{ij}$ s.t. $p+q_{ij} \ge \dfrac{d_{ij}y_{ij}}{2}$ $0 \le r_{ij} \le Mx_{ij}$ $q_{ij}-M(1-x_{ij}) \le r_{ij} \le q_{ij}$ $p, q_{ij}, r_{ij} \ge 0,\ \mathbf{x},\mathbf{y} \in X\times Y.$	B: $2\lvert\mathcal{A}\rvert$ C: $2\lvert\mathcal{A}\rvert+1$ A: $\lvert\mathcal{V}\rvert+4\lvert\mathcal{A}\rvert$ S: $2\lvert\mathcal{A}\rvert+1$
Modified Big-M	$\min\limits_{\substack{\mathbf{x},\mathbf{y}\\ \mathbf{q},\mathbf{r},p}} f(\mathbf{x},\mathbf{y})+p\Gamma+\sum\limits_{(i,j)\in\mathcal{A}} r_{ij}+\sum\limits_{(i,j)\in\mathcal{A}} q_{ij}(1-\gamma_{ij})$ s.t. $p+q_{ij} \ge \dfrac{d_{ij}y_{ij}}{2}$ $r_{ij} \ge \gamma_{ij}-Mx_{ij}$ $p, q_{ij}, r_{ij} \ge 0,\ \mathbf{x},\mathbf{y} \in X\times Y.$	B: $2\lvert\mathcal{A}\rvert$ C: $2\lvert\mathcal{A}\rvert+1$ A: $\lvert\mathcal{V}\rvert+2\lvert\mathcal{A}\rvert$ S: $2\lvert\mathcal{A}\rvert+1$

Table 2.3. Shortest path formulations for the set $\mathcal{U}^{SP}(\mathbf{x})$ (B: Binary, C: Continuous, A: Affine, S: Sign).

and decision constraints. Furthermore, $f(\mathbf{x},\mathbf{y}) = \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} + \sum_{(i,j)\in\mathcal{A}} \bar{d}_{ij}y_{ij}$ denotes

the total cost of reduction and nominal length. Table 2.3 shows that the difference between the Big-M formulation and the other two formulations lies in the number of affine (linear) constraints, as in Table 2.2. We now discuss the numerical experiments.

Experiment 1: Performance Comparison The numerical setup is as follows. We randomly generate points on a 100×100 area and connect them to create a complete graph. The two furthest nodes constitute the source and destination. The final graph is selected after removing 60% of the longest arcs in order to avoid direct connections between the source and destination. The uncertainty budget Γ is set to 2. The cost of reduction $c_{ij} = c$ and the fraction of uncertainty reduced $\gamma_{ij} = \gamma$ are 1.0 and 0.2, respectively. For each size $|\mathcal{V}| = \{50, 75, \ldots, 300\}$, 100 random graphs are generated. These values serve as an illustration of the qualitative comparison of the formulations. In practical applications, they need to be estimated from the economical value of travel time (d_{ij}) relative to the per-trip tax burden for road investments (c_{ij}).

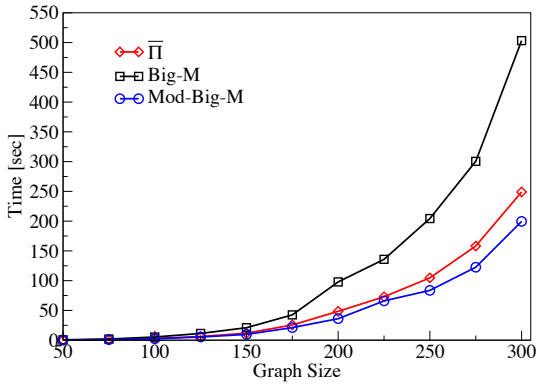


Figure 2.2. Comparison of median solution times of reformulations from Propositions 2.4.1, 2.5.2, and the standard Big-M.

Computational Setup. To solve these problems, we used the Gurobi 7.0 solver on a commercially available computing unit with Intel Core i7 at 3.6 GHz. We reformulated the decision-dependent RO problem as a mixed integer linear program and implemented it using the JuMP library in the Julia programming language v0.6.

The median computation times for different approaches and varying sizes are reported in Figure 2.2. Note that all three methods lead to the same solution. The observations from Figure 2.2 can be summarized as follows.

- The time increases with growing $|\mathcal{V}|$ for all formulations. However, the increase is less steep for the $\overline{\mathbf{\Pi}}$ and the Modified Big-M formulation than for the Big-M formulation.
- The difference between the Big-M and the proposed formulations increases with growing $|\mathcal{V}|$. This highlights the advantage of the $\overline{\mathbf{\Pi}}$ and Modified Big-M formulation for larger graphs.
- The median time of the Modified Big-M formulation is less than that of the $\overline{\mathbf{\Pi}}$-formulation.

Figure 2.2 highlights the benefits of using the proposed formulations to solve such decision-dependent optimization problems. While the performance of the Modified Big-M and $\overline{\mathbf{\Pi}}$ formulations are comparable over a broad range of network sizes, the subproblem in the $\overline{\mathbf{\Pi}}$ reformulation is convex, which can be exploited by cut-generating methods, which may be computationally advantageous. We also solved the $\overline{\mathbf{\Pi}}$ formulation using a cut generation approach (not shown). However, for this application, it converged slowly and required a sizable number of cuts.

We now focus on analyzing how the solution changes as the parameters of the uncertainty set are varied. For this purpose, we introduce additional notation for observable quantities.

Notation for Observables. The number of arcs in the shortest path is n^*, which is a function of the budget Γ and the level of uncertainty reduction γ. These parameters create three scenarios:

(i) *nominal* case, where no uncertainty is present, $n^*(\Gamma = 0, \gamma = 0)$;

(ii) *standard robust* case with no decision dependence, $n^*(\Gamma > 0, \gamma = 0)$; and

(iii) *decision-dependent robust* case with uncertainty reduction $n^*(\Gamma > 0, \gamma > 0)$, in which case $\tilde{n}$ is the number of arcs whose uncertainty was reduced.

We also follow this notation for the optimal objective value z^*. Consequently, the difference $\big(z^*(\Gamma > 0, \gamma = 0) - z^*(\Gamma = 0, \gamma = 0)\big)$ constitutes the *price of robustness*, whereas the difference $\big(z^*(\Gamma > 0, \gamma = 0) - z^*(\Gamma > 0, \gamma > 0)\big)$ constitutes the *benefit of interaction.*

There are four parameters that govern the effect of interactions with uncertainty: $\gamma, |\mathcal{V}|, c$, and Γ. To evaluate their role and to infer the underlying mechanism, we devise four experiments by tuning across their range. Specifically, by adjusting one parameter while keeping the other three fixed, we explore four orthogonal settings.

In these experiments, the problem (SP) is implemented on randomly generated graphs of $[20 - 50]$ nodes. This size is comparable to moderately sized transportation networks [78]. For each size, 2000 graphs are generated in a manner similar to the previous experiment. We maintain these parameter values throughout the following experiments, except in those where their change is probed. In the following, we discuss the four experiments.

Experiment 2: Uncertainty Reduction. We compare z^*, when reduction is permitted ($\gamma > 0$) or not ($\gamma = 0$). Figure 2.3a shows that $\gamma > 0$ reduces z^* (shorter paths), which is independent of $|\mathcal{V}|$. The inset of Figure 2.3a is a magnification, displaying marginal fluctuations that stem from the random nature of graphs.

Experiment 3: Graph Size. We observe that not all arcs in the shortest path experience uncertainty reduction ($\tilde{n} < n^*(\Gamma > 0, \gamma > 0)$), independent of $|\mathcal{V}|$. This is attributed to the non-zero c. We also observe that z^* is independent of $|\mathcal{V}|$, which can be explained by the fact that $|\mathcal{V}|$ only increases from $20 - 50$ and $n^*(\Gamma > 0, \gamma > 0)$ does not change sizably over this range as such the effect on z^* is undetectable. We expect n^* and z^* to increase measurably when $|\mathcal{V}|$ varies by a few orders of magnitude. Larger experiments come at a significant computational burden and are outside the scope of this study.

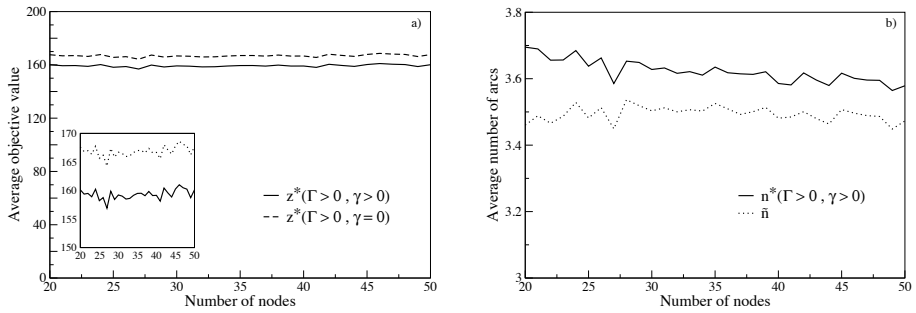


Figure 2.3. Dependence on graph size $|\mathcal{V}|$ for: a) average objective function and b) average number of arcs. The inset is a magnification.

Figure 2.3b illustrates the average $n^*(\Gamma > 0, \gamma > 0)$ and the average $\tilde{n}$ for varying $|\mathcal{V}|$. We also observe a slight downward trend of $n^*(\Gamma > 0, \gamma > 0)$ with increasing $|\mathcal{V}|$. This is because the connectivity within a graph increases with $|\mathcal{V}|$ as the number of arcs grows faster than the number of nodes, because in the experimental setup, only a fixed fraction of arcs are removed.

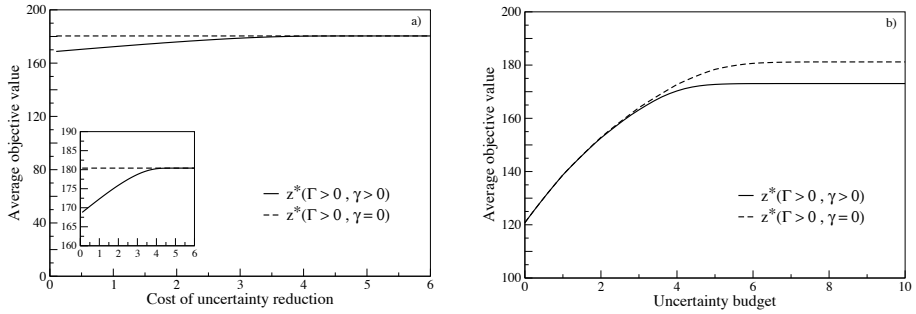


Figure 2.4. Average objective value as a function of: a) cost of uncertainty reduction c and b) maximum uncertainty Γ. The graph consists of $|\mathcal{V}| = 30$ nodes.

Experiment 4: Cost of Uncertainty Reduction. The reduction cost c determines the trade-off between accepting the uncertainty level and its reduction. It can be expected that an increasing c marginalizes the benefits of reducing uncertainty. This means that for a sufficiently low c, uncertainty can be reduced in every arc in the shortest path. On the other hand, for high c, the opposite is true. Figure 2.4a ($|\mathcal{V}| = 30$ and $\Gamma = 12$) shows that for $c \leq 4$, the average z can be decreased. However, for large c, the high cost of reduction makes it disadvantageous to reduce uncertainty. The price of robustness (difference between the dotted line in Figure 2.4a and $z^*(\Gamma = 0, \gamma = 0)$ in Figure 2.4b) is constant w.r.t. γ but changes with Γ. On the other hand, the benefit of interaction decreases with increase in c, as can be observed in Figure 2.5a. Note that the maximum benefit of interaction is calculated by assuming uncertainty is reduced on all the arcs in the shortest path, at zero cost ($c = 0$).

Experiment 5: Uncertainty Budget. Γ governs the number of arcs that can be affected by uncertainty. Figure 2.4b shows that z^* increases gradually with Γ until it reaches the level of the corresponding shortest path length affected by the relative uncertainty $(1 + \frac{1}{2})$

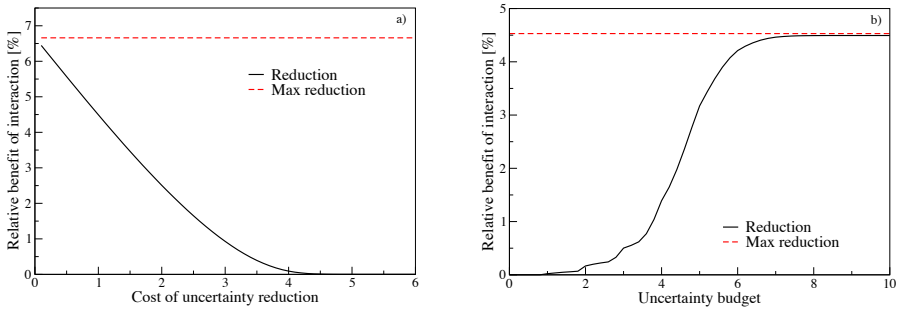


Figure 2.5. Average relative benefit of interaction as a function of: a) cost of uncertainty reduction c and b) maximum uncertainty Γ. The graph consists of $|\mathcal{V}| = 30$ nodes.

and plateaus thereafter. This is because increasing Γ beyond a certain point does not have any effect on n^*, since all the arcs in the path are already uncertain and additional budget remains untapped. Consequently, the price of robustness increases with Γ and plateaus beyond a certain Γ (not shown). An analogous behavior can be observed for the benefit of interaction, as shown in Figure 2.5b. The maximum benefit is achieved at $c = 0$.

Figure 2.6a displays how the average n^* changes with Γ for the different settings. Note that the values of uncertainty are relative to the nominal arc length. This provides an upper bound on the maximum objective value, i.e., when every arc in the shortest path (contributing to n^*) is affected by the uncertainty. At $\Gamma = 0$, we observe $n^*(\Gamma = 0, \gamma = 0)$, and $\tilde{n} = 0$. As Γ increases, it turns beneficial to choose more but shorter arcs, hence, the average $n^*(\Gamma > 0, \gamma = 0)$ initially increases and reaches a maximum at $\Gamma \approx n^*(\Gamma = 0, \gamma = 0)$. As Γ grows even further, the standard robust solution $n^*(\Gamma > 0, \gamma = 0)$ decreases and plateaus at the same level as $n^*(\Gamma = 0, \gamma = 0)$. When $\gamma > 0$, we observe that an increasing $\Gamma \geq 0$ permits more uncertain arc lengths to be reduced ($\tilde{n} \geq 0$) to a maximum

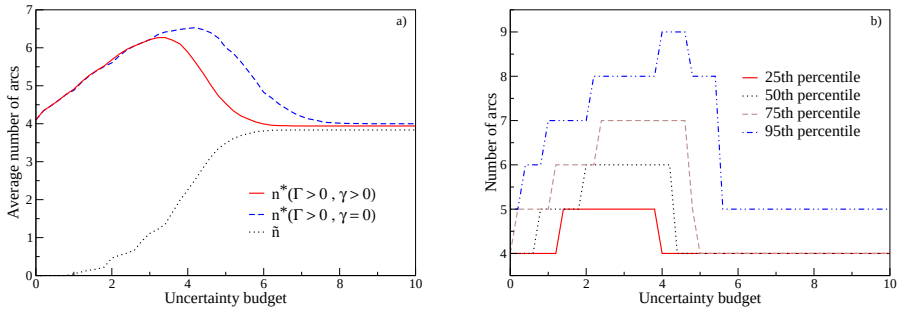


Figure 2.6. The dependence on budget of uncertainty Γ for: a) average number of arcs and b) their distribution. The graph consists of $|\mathcal{V}| = 30$ nodes and uncertainty reduction is permitted.

of $\tilde{n} \leq n^*(\Gamma = 0, \gamma = 0)$. Since some of the arc uncertainty can be reduced, the peak of $n^*(\Gamma > 0, \gamma > 0)$ occurs at a lower budget than when no reduction is allowed, as seen in Figure 2.6a. Note that for small Γ, in order to cope with uncertainty, the optimal solution minimizes the length of each individual arc so that the impact of the uncertainty is minimized.

To further support this observation, Figure 2.6b displays the distribution of the number of arcs using different percentiles of $n^*(\Gamma > 0, \gamma > 0)$ (corresponding to Figure 2.6a). Here, we observe that as Γ increases, the distribution of $n^*(\Gamma > 0, \gamma > 0)$ skews towards larger number of arcs (the gaps between the percentiles increase). This means that the optimal solution becomes more diversified. Specifically, the model selects a path consisting of some certain and some uncertain arcs, with a subset of the latter experiencing uncertainty reduction. This continues until the saturation point (here $\Gamma \approx 4$) because beyond a certain budget, diversification of paths becomes redundant. At this point, the shortest path is chosen exclusively amongst uncertain arcs, almost all of which experience uncertainty reduction (since $\Gamma > n^*(\Gamma = 0, \gamma = 0)$).

Experiment 6: Comparison to SO. This experiment evaluates the average and worst case performance of the robust DDU solutions and compares them to a similar SO problem. The SO formulation is given by

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} + \mathbb{E}_{\mathbb{P}(\mathbf{x})}\left[\sum_{(i,j)\in\mathcal{A}} d_{ij}(\boldsymbol{\xi})y_{ij}\right] \\ \text{s.t.} \quad & \mathbf{y} \in Y \\ & \mathbf{x} \in \{0,1\}^{|\mathcal{A}|}, \end{aligned}$$

with the uncertainty set

$$\boldsymbol{\xi} \in \mathcal{U}^{SSP}(\mathbf{x}) = \times_{i,j\in\mathcal{A}}[0, 1-\gamma x_{ij}].$$

The distribution $\mathbb{P}(\mathbf{x})$ is the uniform distribution over the support $\mathcal{U}^{SSP}(\mathbf{x})$. The explicit expression for the expected value is displayed in Nohadani and Sharma [82]. The average performance is evaluated by randomly generating the uncertain component ξ_{ij} (from $[0,1]$ for unreduced arcs and $[0,1-\gamma x_{ij}]$ for reduced arcs) and implementing the existing robust and stochastic solutions for these randomly generated arc costs. The following solutions are evaluated: (i) RO: Robust solution for $\gamma = 0$. (ii) RO-DDU: Robust solution for $\gamma > 0$. (iii) SO: Stochastic solution for $\gamma = 0$. (iv) SO-DDU: Stochastic solution for $\gamma > 0$. The suffix of the average performances is "-A" and of the worst-case performances "-W."

Figure 2.7a shows that the average objective of SO is less than the average RO objective. This is because RO optimizes the worst case instead of the average performance as in SO. However, analogously in Figure 2.7b, RO-W is significantly less than SO-W. The

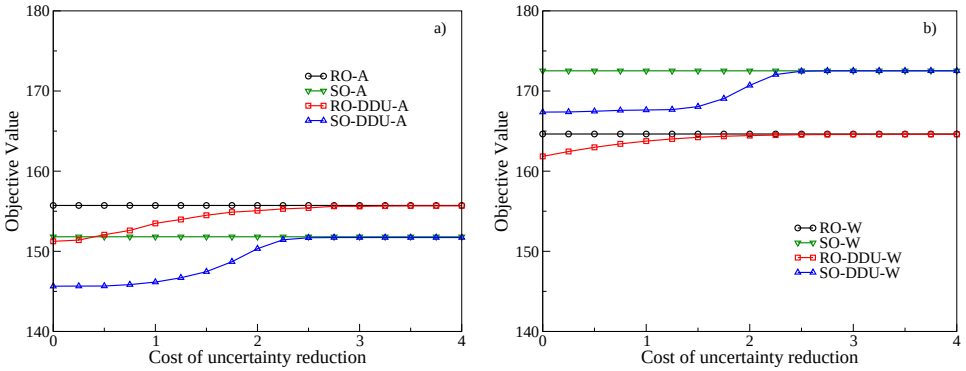


Figure 2.7. Comparison of RO and SO formulations: a) average and b) worst-case objective value.

same applies to the decision-dependent counterparts for both cases. As can be expected, the objective values increase with c until it is no longer beneficial to reduce the uncertainty, i.e., the objective value of the RO-DDU solution increases until it matches that of the RO solution. The same holds true for the SO-DDU and SO solutions.

In summary, the $\overline{\mathbf{\Pi}}$-formulation and the Modified Big-M formulation perform considerably better than the standard Big-M formulation and their benefits increase with graph size. The worst-case cost for the shortest path can be improved by proactively reducing the uncertainty on a subset of arcs. As the budget of uncertainty grows, these benefits improve but plateau beyond a certain level. At the same time, the cost of reduction curbs these benefits. The RO-DDU problem performs better than SO-DDU for the worst-case scenario. As expected, this benefit comes at the price of the average cost. This numerical study provides an overview of the impact of different formulations, probes various model parameters, and highlights the power of the proactive uncertainty control for both the worst-case and average performance.

2.7. Conclusion

In this chapter, we present a novel optimization approach for solving problems with decision-dependent uncertainties. We show that for general polyhedral sets, such problems are, even in basic cases, NP-complete. To alleviate this, we introduce a class of uncertainty sets whose upper bounds are affected by decisions. They enable more realistic modeling of a broad range of applications and extend RO beyond the currently used exogenous sets. We provide reformulations that have considerably fewer constraints compared to standard linearization techniques, allowing for faster computations. Our approach should be viewed as one option among many to model decision dependence while maintaining computational advantages. The induced convexity of the sub-problem in the proposed reformulation reveals a path forward to use advanced cut generating algorithms. We believe that finding new and appropriate conditions on sets will further improve the quality of the reformulations.

In addition, this chapter provides an alternative way of addressing one of the criticisms of RO approaches, namely overly conservative solutions. The description via decision-dependent sets enables mitigation of this issue by exercising proactive control on uncertainties. This setting offers an immediate way to manage the tradeoff between conservatism and optimality. Finally, novel cutting plane methods have instrumentally enhanced solution times and we envision decision-dependent sets to solidify the tradeoff between computation and optimality by inducing beneficial cuts.

In the next chapter, we extend our results on decision-dependent uncertainty sets to allow for uncertain influence of the decisions on the sets. We further elaborate the

discussion on the use of cutting place methods for decision-dependent sets in the context of a unit commitment problem.

CHAPTER 3

Decision Dependent Uncertainty with uncertain Influence

3.1. Introduction

The previous chapter discussed robust optimization problems with decision dependent uncertainty. In these problems, a decision variable affects the uncertainty set by modifying the parameters of the set. In the example of the set

$$\mathcal{U}(\mathbf{x}, \mathbf{W}) = \{\boldsymbol{\xi} \mid \mathbf{D}\boldsymbol{\xi} \leq \mathbf{d},\ \mathbf{0} \leq \boldsymbol{\xi} \leq \mathbf{v} + \mathbf{W}(\mathbf{e} - \mathbf{x})\},$$

where $\mathbf{d}, \mathbf{D}$ and $\mathbf{v}$ are parameters of the set and $\mathbf{e}$ is a vector of all ones, the decision variable $\mathbf{x} \in X \subseteq \{0,1\}^n$, changes the set by modifying the upper bound on the uncertain parameter $\boldsymbol{\xi}$. Thus, the maximum possible value of $\boldsymbol{\xi}$ changes, depending on $\mathbf{x}$. The exact value of the upper bound plays a key role in the optimal solution of the optimization problem. As such, evaluating the exact influence of decisions on the uncertainty set is important. This influence, however, depends on the values of the parameter $\mathbf{W}$, which may not be known a priori and be uncertain as well. For example, in the shortest path problem discussed in Chapter 2, the amount of reduction in the worst-case path length may be unknown. This notion of uncertain influence can naturally exist when decision dependent uncertainty involves scenario trees [46, 47, 58]. In these problems, the decision can choose a branch of the scenario tree but exact impact of this choice on the future may be uncertain as it may not be known what information is going to be revealed by that

choice. In this chapter, we tackle this problem by incorporating the uncertainty present in the value of the *set influencing parameter*, $\mathbf{W}$, into the optimization problem.

3.2. Model

Chapter 2 focused on robust linear optimization problems with decision dependent uncertainty. It introduced a special class of uncertainty sets that led to a direct convex reformulation of the constraints. Here, we further refine the same class of sets to illustrate the incorporation of uncertainty in the set influencing parameter into the optimization problem. Consider the set $\mathcal{U}(\mathbf{x}, \mathbf{W}) = \{\boldsymbol{\xi} \mid \mathbf{D}\boldsymbol{\xi} \leq \mathbf{d},\ \mathbf{0} \leq \boldsymbol{\xi} \leq \mathbf{v} + \mathbf{W}(\mathbf{e} - \mathbf{x})\}$. The parameter $\mathbf{W}$ controls the influence of the decision $\mathbf{x}$ on the set. For any fixed value of $\mathbf{W}$ we can write the optimization problem

$$\min_{\mathbf{x} \in X} \left\{ \mathbf{c}^\top \mathbf{x} + \max_{\boldsymbol{\xi} \in \mathcal{U}(\mathbf{x}, \mathbf{W})} \min_{\mathbf{y} \in Y(\mathbf{x}, \boldsymbol{\xi})} \mathbf{r}^\top \mathbf{y} \right\}. \tag{3.1}$$

Here, $\mathbf{x}$ and $\mathbf{y}$ are the first and second stage decisions and $\boldsymbol{\xi}$ is the uncertain component. $Y(\mathbf{x}, \boldsymbol{\xi})$ is the feasibility set for the second stage problem. We now express the dependence of $\mathbf{U}$ on $\mathbf{W}$ explicitly.

When the parameter $\mathbf{W}$ is uncertain, its uncertainty can be described via sets or via distributions. Examples are

- Interval Sets:

$$\mathbf{W} \in \mathcal{W} = \{\mathbf{W} : \underline{\mathbf{W}} \leq \mathbf{W} \leq \overline{\mathbf{W}}\},$$

- Finite distribution:

$$\mathbf{W} \sim \mathcal{W} : \mathbb{P}[\mathbf{W} = \mathbf{W}_s] = \pi_s \ \ \forall s = 1, \ldots, N_s.$$

Depending on choice of a set based approach or a distribution based approach, we formulate the resulting problem differently, as we will describe in the following.

3.2.1. Robust formulation for set based information

In most cases, information about the parameter $\mathbf{W}$ may be limited. This is because we may not be able to observe this reduction directly and may have to infer its value from multiple observations of the uncertainty $\boldsymbol{\xi}$. In these settings, robust optimization is an appropriate approach to incorporating any uncertainty in $\mathbf{W}$ into the optimization problem. Let the uncertain parameter $\mathbf{W}$ lie in an uncertainty set $\mathcal{W}$. Then we can express a robust reformulation of the problem (3.1) as,

$$\min_{\mathbf{x}\in X}\left\{\mathbf{c}^\top\mathbf{x}+\max_{\mathbf{W}\in\mathcal{W}}\left[\max_{\boldsymbol{\xi}\in\mathcal{U}(\mathbf{x},\mathbf{W})}\min_{\mathbf{y}\in Y(\mathbf{x},\boldsymbol{\xi})}\mathbf{r}^\top\mathbf{y}\right]\right\}. \tag{3.2}$$

In this problem we consider the worst case over elements $\mathbf{W}$ which are contained in the set $\mathcal{W}$. This problem can be reformulated as following

Proposition 3.2.1. *The problem* (3.2) *can be reformulated as*

$$\min_{\mathbf{x}\in X}\left\{\mathbf{c}^\top\mathbf{x}+\max_{\boldsymbol{\xi}\in\mathcal{U}'(\mathbf{x})}\min_{\mathbf{y}\in Y(\mathbf{x},\boldsymbol{\xi})}\mathbf{r}^\top\mathbf{y}\right\}, \tag{3.3}$$

where $\mathcal{U}'(\mathbf{x})=\bigcup_{\mathbf{W}\in\mathcal{W}}\mathcal{U}(\mathbf{x},\mathbf{W})=\{\boldsymbol{\xi}\mid\mathbf{D}\boldsymbol{\xi}\leq\mathbf{d},\ \mathbf{0}\leq\boldsymbol{\xi}\leq\mathbf{v}+\mathbf{W}(\mathbf{e}-\mathbf{x}),\ \mathbf{W}\in\mathcal{W}\}.$

Proof. It is sufficient to prove that

$$\bigcup_{\mathbf{W}\in\mathcal{W}}\mathcal{U}(\mathbf{x},\mathbf{W})=\{\boldsymbol{\xi}\mid\mathbf{D}\boldsymbol{\xi}\leq\mathbf{d},\ \mathbf{0}\leq\boldsymbol{\xi}\leq\mathbf{v}+\mathbf{W}(\mathbf{e}-\mathbf{x}),\ \mathbf{W}\in\mathcal{W}\}.$$

Both inclusions hold by the definition of the union. □

When the uncertainty set $\mathcal{W}$ is an interval uncertainty set, then the following proposition gives an exact characterization of the worst case.

Proposition 3.2.2. *If the set* $\mathcal{W} = \{\mathbf{W} : \underline{\mathbf{W}} \leq \mathbf{W} \leq \overline{\mathbf{W}}\}$*, then problem* (3.2) *can be reformulated as*

$$\min_{\mathbf{x}\in X} \left\{ \mathbf{c}^\top \mathbf{x} + \max_{\boldsymbol{\xi}\in\mathcal{U}(\mathbf{x},\overline{\mathbf{W}})} \min_{\mathbf{y}\in Y(\mathbf{x},\boldsymbol{\xi})} \mathbf{r}^\top \mathbf{y} \right\}. \tag{3.4}$$

Proof. It is sufficient to prove that

$$\bigcup_{\mathbf{W}\in\mathcal{W}} \mathcal{U}(\mathbf{x}, \mathbf{W}) = \mathcal{U}(\mathbf{x}, \overline{\mathbf{W}}).$$

By the definition of $\mathcal{W}$, we have the following inclusion for any $\mathbf{x}$,

$$\begin{aligned} &\mathcal{U}(\mathbf{x}, \mathbf{W}) \subseteq \mathcal{U}(\mathbf{x}, \overline{\mathbf{W}}) \ \ \forall \mathbf{W} \in \mathcal{W}, \\ \implies &\bigcup_{\mathbf{W}\in\mathcal{W}} \mathcal{U}(\mathbf{x}, \mathbf{W}) \subseteq \mathcal{U}(\mathbf{x}, \overline{\mathbf{W}}). \end{aligned}$$

The opposite inclusion holds true because $\overline{\mathbf{W}} \in \mathcal{W}$. This proves our result. □

Note that the problems (3.2) and (3.4) are adjustable robust optimization problems. They can solved using a column and constraint generation approach [105] or approximated using decision rules.

The robust formulation requires only limited knowledge about $\mathbf{W}$ and thus can be tractably used in many situations. However, it comes with the caveat of focusing on the worst case and cannot capture situations in which the decision is on the information gathering as needed to assemble and motivate the distribution on data. To address this,

we also develop an optimization model that assumes distributional knowledge about the parameter $\mathbf{W}$.

3.2.2. Stochastic Formulation for distribution information

When some knowledge about the distribution of the parameter $\mathbf{W}$ is available, we can evaluate the expected value instead of the worst case. The resulting problem can be expressed as

$$\min_{\mathbf{x}\in X}\left\{\mathbf{c}^\top\mathbf{x}+\mathbb{E}_{\mathbf{W}}\left[\max_{\boldsymbol{\xi}\in\mathcal{U}(\mathbf{x},\mathbf{W})}\min_{\mathbf{y}\in Y(\mathbf{x},\boldsymbol{\xi})}\mathbf{r}^\top\mathbf{y}\right]\right\}. \tag{3.5}$$

Here, we assume that the distribution of $\mathbf{W}$ is given by a finite set $\mathcal{N}_s$ of scenarios $\mathbf{W}_s$ with associated probabilities π_s. The resulting optimization problem can be expressed as

$$\min_{\mathbf{x}\in X}\left\{\mathbf{c}^\top\mathbf{x}+\sum_{s\in\mathcal{N}_s}\pi_s\left[\max_{\boldsymbol{\xi}\in\mathcal{U}(\mathbf{x},\mathbf{W}_s)}\min_{\mathbf{y}\in Y(\mathbf{x},\boldsymbol{\xi})}\mathbf{r}^\top\mathbf{y}\right]\right\}. \tag{3.6}$$

Problem (3.6) is an adjustable robust optimization problem with $|\mathcal{N}_s|$ second-stage scenarios. We can solve this problem using an affine policy approach in which each second stage has its own affine policy that may be different from those of other second stages. This difference arises due to the dependence of these policies on the scenario s. Depending on the application, this dependence has an impact on the optimization problem. For example, in an information gathering problem, we can observe the gathered data and use a policy depending on it. However, in other applications we may not directly observe a realization of $\mathbf{W}$ and instead only observe $\boldsymbol{\xi}$, for example the shortest path problem discussed in chapter 2. In this case, we may not be able to distinguish between the different scenarios and corresponding policies. We do not address this case in this thesis.

In order to illustrate the concept of uncertain influence in a concrete fashion, we conduct a numerical experiment on an electricity unit commitment problem with uncertain load, which is an extension of the work by Lorca et al. [74].

3.3. Unit Commitment Problem

Electric power systems need protection against multiple sources of uncertainty such as load uncertainty [74], transmission line security [100], and power generation uncertainty [95]. Unit commitment problems model such power systems and optimize for a power generation schedule that ensures all loads are met while minimizing cost. It aims to protect against all sources of uncertainty, while achieving its objective. Both stochastic [88, 93] and robust optimization [74] have been used to model uncertainty in unit commitment context problems. SO uses distributional knowledge about the various uncertainty sources such as load, power generation etc. to model the problem. It primarily focuses on multistage problems with finitely many scenarios and leverages the structure of the problem for column or constraint generation procedures to improve computation. RO captures the uncertainty through sets and focuses on using decision rules to reduce computational difficulty. Though, cut generation procedures are still required by RO problem due to the size of the unit commitment problem, the use of decision rules allows us to avoid the use of non-anticipativity constraints as they can be naturally built-in. The primary benefit of using RO models is the potential for solving larger problems due to fewer constraints, however, it comes at the cost of focusing on the worst-case scenarios instead of more realistically occurring ones.

Unit commitment problems are primarily modeled as two-stage or multistage optimization problems as the power generation depends on the realized load and the startup and shutdown time of generators.

We focus on a multistage RO model for the unit commitment problem as used in [74].

3.3.1. Model

Consider the following model of a two-stage robust unit commitment problem:

$$
\begin{aligned}
&\min_{\mathbf{x},\mathbf{z},\mathbf{u},\mathbf{v}} \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}_g}(G_i z_i^t + S_i u_i^t) + \sum_{t\in\mathcal{T}} c_j^t x_j^t + \max_{\boldsymbol{\xi}\in\mathcal{U}(\mathbf{x})} Q(\mathbf{x},\mathbf{z},\mathbf{u},\mathbf{v};\boldsymbol{\xi}) \\
&\text{s.t. } z_i^t - z_i^{t-1} = u_i^t - v_i^t \qquad \forall i\in\mathcal{N}_g,\ \forall t\in\mathcal{T} \quad \text{start and shutdown constraints} \\
&\sum_{\tau=t}^{t+UT_i-1} z_i^\tau \geq UT_i u_i^t \qquad \forall i\in\mathcal{N}_g,\ \forall t\in\mathcal{T} \qquad \text{up time constraints} \\
&\sum_{\tau=t}^{T}(z_i^\tau - u_i^t) \geq 0 \qquad \forall i\in\mathcal{N}_g,\ \forall t\in\mathcal{T} \\
&\sum_{\tau=t}^{t+DT_i-1}(1-z_i^\tau) \geq DT_i v_i^t \ \ \forall i\in\mathcal{N}_g,\ \forall t\in\mathcal{T} \quad \text{down time constraints} \\
&\sum_{\tau=t}^{T}(1-z_i^\tau - v_i^t) \geq 0 \qquad \forall i\in\mathcal{N}_g,\ \forall t\in\mathcal{T} \\
&x_i^t, z_i^t, u_i^t, v_i^t \in \{0,1\} \qquad \forall i\in\mathcal{N}_g,\ \forall t\in\mathcal{T}.
\end{aligned}
\tag{UC}
$$

The second stage problem

$$
\begin{aligned}
&Q(\mathbf{x}, \mathbf{z}, \mathbf{u}, \mathbf{v}; \boldsymbol{\xi}) = \\
&\min_{\mathbf{y}} \sum_{t=1}^{T} \sum_{i \in N_g} r_i y_i^t \\
&\text{s.t. } p_i^{min} z_i^t \leq y_i^t \leq p_i^{max} z_i^t && \forall i \in \mathcal{N}_g,\ \forall t \in \mathcal{T} && \text{bounds on power generated} \\
&\quad - RD_i z_i^t - SD_i v_i^t \leq y_i^t - y_i^{t-1} \leq RU_i^{t-1} z_i^{t-1} + SU_i u_i^t && && \text{ramping constraints} \\
&\quad - f_l^{max} \leq \boldsymbol{\alpha}_l^\top (\mathbf{B}^p \mathbf{y}^t - \mathbf{B}^d \boldsymbol{\xi}^t) \leq f_l^{max} && \forall l \in \mathcal{N}_l,\ \forall t \in \mathcal{T} && \text{bounds on line flow} \\
&\quad \sum_{i \in \mathcal{N}_g} y_i^t = \sum_{j \in \mathcal{N}_d} \xi_j^t && \forall t \in \mathcal{T}. && \text{demand satisfaction}
\end{aligned}
$$

Here, $\mathcal{N}_g, \mathcal{N}_d, \mathcal{N}_l$ denotes the set of generators, nodes with load, and transmission lines. The first stage decisions z_i^t, u_i^t, and v_i^t correspond to on/off, startup, and shutdown decisions respectively, while the second stage decision y_i^t is the amount of power generated from the available generators i. The variable x_j^t controls whether to reduce the uncertainty in the load in period t in node j. The other parameters $G_i, S_i, c_j^t, UT_i, DT_i$ represent the fixed running cost, start up cost, uncertainty reduction cost, minimum up time, and minimum down time, respectively.

For the second stage problem, the parameter $\boldsymbol{\xi}$ represents the uncertain load. From the rest, $r_i, p_i^{min}, p_i^{max}, RD_i, SD_i, RU_i, SU_i$ are the power generation costs, minimum power generated, maximum power generated, ramp down rate, shut down rate, ramp up rate, and start up rates, respectively. For the line flow constraint, f_l represents the maximum power flow through a line, $\boldsymbol{\alpha}_l$ is the transmission line distribution factor, and $\mathbf{B}^p$ and $\mathbf{B}^d$ are the generation and load incidence matrices.

We assume that only the load is subject to uncertainty in this problem. The set capturing this uncertainty is given by $\mathcal{U}(\mathbf{x}) = \mathcal{U}^1(\mathbf{x}^1) \times \mathcal{U}^2(\mathbf{x}^2) \times \cdots \times \mathcal{U}^T(\mathbf{x}^T)$, where

$$\mathcal{U}^t(\mathbf{x}^t) = \left\{ \boldsymbol{\xi}^t \;\middle|\; \sum_{j \in \mathcal{N}_d} \frac{|\xi_j^t - \bar{\xi}_j^t|}{\hat{\xi}_j^t} \leq \Gamma \sqrt{|\mathcal{N}_d|},\;\; \xi_j^t \in [\bar{\xi}_j^t - \Gamma \hat{\xi}_j^t, \bar{\xi}_j^t + (\Gamma - \gamma x_j^t)\hat{\xi}_j^t] \right\}.$$

This set is a cardinality constrained uncertainty set [18] and is inspired by the central limit theorem. The decision x^t allows us to modify the upper bound on the uncertain parameter and reduce the maximum uncertainty in all nodes in the time period t.

Problem (UC) is an adjustable robust optimization problem with decision dependent uncertainty. Therefore, it is difficult to solve directly. In order to allow for computational tractability, we simplify problem (UC) by using affine decision rules

$$y_i^t(\boldsymbol{\xi}^{[t]}) = w_i^t + W_{it} \sum_{j \in \mathcal{N}_d} \xi_j^t \quad \forall i \in \mathcal{N}_g,\ \forall t \in \mathcal{T}. \tag{3.7}$$

With this assumption, we can use the results from [74]. The demand satisfaction constraint can be reformulated as follows.

Proposition 3.3.1 (Lorca et al. [74])**.** *For a full-dimensional uncertainty set, the robust demand satisfaction constraints with affine decision rules of the form* (3.7) *are equivalent to*

$$\sum_{i \in \mathcal{N}_g} w_i^t = 0, \quad \sum_{i \in \mathcal{N}_g} W_{it} = 1 \;\; \forall t.$$

This result extends to decision-dependent uncertainty sets without any changes, if $\gamma < \Gamma$. This assumption holds, if we require each load to have some minimum level of uncertainty. The remaining robust constraints can be reformulated using duality theory. However, due to the large number of constraints in the resulting reformulation, we use a cut generation algorithm to

incorporate the robust constraints into the problem. This cut generation process is possible for decision-dependent uncertainty set by the results presented in chapter 2.

3.4. Cut generation

All the robust constraints of problem (UC), except the line flow constraints, can be expressed in the form $\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{N}_d}\alpha_j^t\xi_j^t \leq h \quad \forall \boldsymbol{\xi} \in \mathcal{U}(\mathbf{x})$, where $\boldsymbol{\alpha}$ is a linear function of the decision variable $\mathbf{y}$. We can write this as

$$\boldsymbol{\alpha}^\top \boldsymbol{\xi} \leq h \;\; \forall \boldsymbol{\xi} \in \mathcal{U}(\mathbf{x}),$$

for the uncertainty set

$$\mathcal{U}^{\overline{\mathbf{\Pi}}}(\mathbf{x}) = \{\boldsymbol{\xi} \mid \mathbf{G}\boldsymbol{\xi} \leq \mathbf{g},\;\; \boldsymbol{\xi} \leq \mathbf{v} + \mathbf{W}(\mathbf{e}-\mathbf{x}),\;\; \boldsymbol{\xi} \geq \mathbf{0}\}.$$

Using the results from Chapter 2, this constraint can be reformulated as

$$\begin{aligned} &\max_{\boldsymbol{z},\boldsymbol{\zeta}} \; (\boldsymbol{\alpha} - \overline{\mathbf{\Pi}}\mathbf{x})^\top \boldsymbol{\chi} + \boldsymbol{\alpha}^\top \boldsymbol{\zeta} \\ &\text{s.t. } \mathbf{G}(\boldsymbol{\chi}+\boldsymbol{\zeta}) \leq \mathbf{g} \\ &\qquad \boldsymbol{\chi} \leq \mathbf{We} \\ &\qquad \boldsymbol{\zeta} \leq \mathbf{v} \\ &\qquad \boldsymbol{\chi}, \boldsymbol{\zeta} \geq \mathbf{0}. \end{aligned} \tag{3.8}$$

We can reformulate the above problem using duality, however, the large number of constraints in (UC) leads to a very large robust problem which is difficult so solve. To avoid this, we develop a cut generation algorithm, which allows us to generate the worst-case scenarios corresponding to each constraint iteratively. For robust constraint, we solve the problem (3.8) and identify a scenario which violates the constraint. This violated constraint is then added to the master problem. Once we have added a violation corresponding to each of the constraints, the master

problem is re-solved. This process is repeated until we are unable to find any violated constraints, which solves the overall problem.

3.4.1. Pre-computed scenarios

The cut generation algorithm allows us to reduce the number of constraints in the problem. However, it still requires the master problems to be solved repeatedly many times as the constraints are generated. This significantly increases the computational difficulty. To alleviate this issue, we can pre-compute the worst-case scenarios for some of the constraints. These pre-computed scenarios can then be added at the initialization of the problem. Consequently, the corresponding constraints do not need to be generated iteratively. For this purpose, we make the following assumption,

Assumption 3.4.1. *The uncertainty influencing variable* $\mathbf{x}$ *does not depend on the nodes* j *with net load.*

This assumption enables us to pre-compute the scenarios and solve realistic size unit commitment problems. In practice, such an assumption implies that any reduction in load uncertainty happens across the board for all users at the specified times.

Consider the problem $\max_{\boldsymbol{\xi}^t \in \mathcal{U}^t(x^t)} \alpha_{it} \sum_{j \in \mathcal{N}_d} \xi_j^t$ where

$$\mathcal{U}^t(x^t) = \left\{ \sum_{j \in N} \frac{|\xi_j^t - \bar{\xi}_j^t|}{\hat{\xi}_j^t} \leq \Gamma\sqrt{N}, \;\; \xi_j^t \in [\bar{\xi}_j^t - \Gamma\hat{\xi}_j^t, \bar{\xi}_j^t + (\Gamma - \gamma x_i^t)\hat{\xi}_j^t] \right\}.$$

Using the results from chapter 2, this problem can be expressed as

$$
\begin{aligned}
&\max_{\boldsymbol{\chi},\boldsymbol{\zeta}} \sum_{j=1}^{N_d}((\alpha_{it} - Mx^t)\chi_{jt} + \alpha_{it}\zeta_{jt}) \\
&\text{s.t.} \sum_{j=1}^{N_d} \frac{|\chi_{jt} + \zeta_{jt} - \Gamma\hat{d}_{jt}|}{\hat{d}_{jt}} \leq \Gamma\sqrt{N_d} \\
&0 \leq \zeta_{jt} \leq 2\Gamma\hat{d}_{jt} - \hat{d}_{jt}\gamma_{jt} \;\; \forall j \\
&0 \leq \chi_{jt} \leq \hat{d}_{jt}\gamma_{jt} \;\; \forall j.
\end{aligned} \tag{3.9}
$$

We have the following proposition for the solution to this problem.

Proposition 3.4.1. *For any given α_{it} and x^t, the optimal solution to problem* (3.9) *is*

$$
\begin{aligned}
&(\alpha_{it} \geq 0, x^t = 1) : (\boldsymbol{\chi}^t, \boldsymbol{\zeta}^t) = \arg\max_{\boldsymbol{\chi},\boldsymbol{\zeta}} \sum_{j=1}^{N_d}(-M\chi_{jt} + \zeta_{jt}) \\
&(\alpha_{it} \geq 0, x^t = 0) : (\boldsymbol{\chi}^t, \boldsymbol{\zeta}^t) = \arg\max_{\boldsymbol{\chi},\boldsymbol{\zeta}} \sum_{j=1}^{N_d}(\chi_{jt} + \zeta_{jt}) \\
&(\alpha_{it} < 0, x^t = 1) : (\boldsymbol{\chi}^t, \boldsymbol{\zeta}^t) = \arg\max_{\boldsymbol{\chi},\boldsymbol{\zeta}} \sum_{j=1}^{N_d}(-M\chi_{jt} - \zeta_{jt}) \\
&(\alpha_{it} < 0, x^t = 0) : (\boldsymbol{\chi}^t, \boldsymbol{\zeta}^t) = \arg\max_{\boldsymbol{\chi},\boldsymbol{\zeta}} \sum_{j=1}^{N_d}(-\chi_{jt} - \zeta_{jt}).
\end{aligned} \tag{3.10}
$$

Proof. Case 1: If $x^t = 0$, then the objective of (3.9) is $\max_{\boldsymbol{\chi},\boldsymbol{\zeta}\in\mathcal{U}^t} \alpha_{it} \sum_{j=1}^{N_d}(\chi_{jt} + \zeta_{jt})$. The solution of this problem then only depends on the sign of α_{it} and not on its value.

Case 2: If $x^t = 1$, then the objective function becomes $\max_{\boldsymbol{\chi},\boldsymbol{\zeta}\in\mathcal{U}^t} \sum_{j=1}^{N_d}((\alpha_{it} - M)\chi_{jt} + \alpha_{it}\zeta_{jt})$.

If $\alpha_{it} > 0$, then the objective is $\alpha_{it} \max_{\boldsymbol{\chi},\boldsymbol{\zeta}\in\mathcal{U}^t} \sum_{j=1}^{N_d}((1 - \frac{M}{\alpha_{it}})\chi_{jt} + \zeta_{jt})$.

Let $M' = \frac{M}{\alpha_{it}}$. Since M is large, this means that $M' >> 1$, and we can write the objective as $\alpha_{it} \max_{\boldsymbol{\chi},\boldsymbol{\zeta}\in\mathcal{U}^t} \sum_{j=1}^{N_d}(-M'\chi_{jt} + \zeta_{jt})$. For the case of $\alpha_{it} < 0$, we can arrive at the result similarly. This completes the proof. □

Using the proposition (3.4.1), we can pre-compute the worst-case scenarios for any of the constraints which can be expressed in the form $\max_{\boldsymbol{\xi}^t \in \mathcal{U}^t(x^t)} \alpha_{it} \sum_{j \in \mathcal{N}_d} \xi_j^t \leq h_{it} \ \ \forall i \in \mathcal{N}_g, \ \forall t \in \mathcal{T}$. This corresponds to the structure of the constraints which bound the power generated. For the ramping constraints and the line flow constraints we use the standard cut generation process. We now conduct numerical experiments on the final problem.

3.4.2. Numerical Experiments

We illustrate the benefits of the decision dependent approach as well as the impact of uncertain reduction on a unit commitment problem. Specifically, we focus on a reduced version of the IEEE-118 bus problem [1] with only 15 buses.

3.4.2.1. Setup. The reduced problem has $|\mathcal{N}_g| = 9$ generators, $|\mathcal{N}_d| = 11$ loads, and focuses on a period of $|\mathcal{T}| = 12$ hours. The values of the other parameters are $\Gamma = 2$, cost of reduction $r^t = 10$/load/time and $\hat{d}_j = 0.1\bar{d}_j$. The average loads are taken from the original 118 bus problem.

Computational Setup. The master problem is modeled as a MILP while the sub problem is solved as a linear program. The optimization problem is modeled using JuMP library on the Julia programming language v1.0 and solved with the Gurobi v8 solver. The experiments are run on a machine with an Intel Core i7 processor with 32 GB RAM.

3.4.2.2. Results. From Figure 3.1, it can be observed that as the price of reducing the uncertainty increases, the benefit of reducing it decreases along with the amount of reduction. This indicates that the price that needs to be paid plays a key role in determining whether it is worth reducing the uncertainty. The figures also show the key role of the parameter γ, i.e., the amount of reduction in the uncertainty. The benefit of reducing the uncertainty as well as the load reduced changes significantly with the value of the parameter γ.

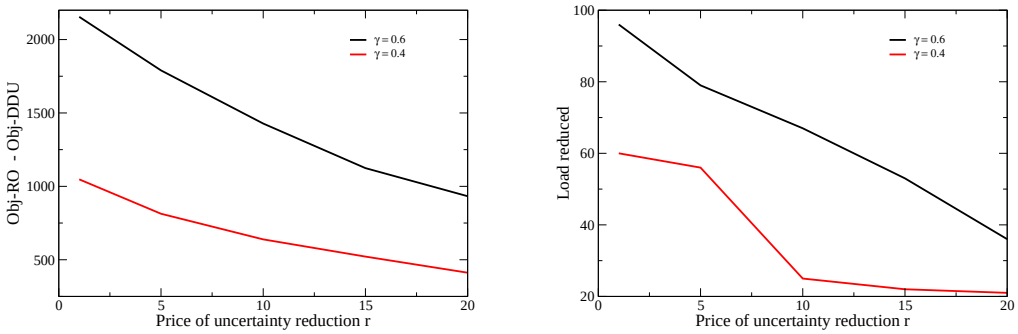


Figure 3.1. Benefit of DDU (left) and Total load reduced (right)

Model	γ	**Reduced**	**Obj**	**Change**
Nominal			103926	
Robust	0		115516	11.2%
DDU Fixed	0.5	Yes	114557	10.2%
DDU Robust 1	[0.4, 0.6]	Yes	115216	10.9%
DDU Robust 2	[0.3, 0.7]	No	115516	11.2%
DDU Stoch. 1	{0.4, 0.5, 0.6}	Yes	114371	10.0%
DDU Stoch. 2	{0.3, 0.5, 0.7}	Yes	114083	9.8%

Table 3.1. Comparison of uncertainty models

Table 3.1 shows the results for multiple uncertainty models for the unit commitment problem. We can observe that robustness comes at a price as it increases the objective function. However, this cost can be reduced by using decision dependent uncertainty. The benefit, however, depends on the value of γ. For example, it is worthwhile to reduce the uncertainty for the model DDU Robust 1, but not for DDU Robust 2. This is because worst value of γ is 0.3 for the DDU Robust 2, which is lower than 0.4 for DDU Robust 1. However, the situation is different for the stochastic

model, because it focuses on the expected value instead of the worst-case as such the DDU Stoch. 2 model still reduces the uncertainty even though it has a support with bounds similar to DDU Robust 2. However, the computational challenge of solving the Stochastic models is much higher than that of the Robust models. Overall, we can observe that the presence of uncertainty in the amount of reduction γ can have significant impact and it is necessary to incorporate it into the optimization problem.

3.5. Conclusion

In this chapter, we extended decision-dependent uncertainty sets to also allow for uncertainty in the influence of the decisions on the set. We discussed methods to incorporate this uncertainty into the overall optimization problem by modeling the uncertainty in the influence as set based or distribution based. This led to robust and stochastic optimization problems. The robust problem focuses on worst cases, while the stochastic problem addresses on the expected value. This increases the difficulty of solving the stochastic problems even though they may be more realistic. In the numerical experiments on the unit commitment problem, we developed a cut generation algorithm and provided results to pre-compute the worst-case scenarios for some of the constraints, which improves the computability of the problem. The experiments illustrated the benefits of incorporating any uncertainty in the influence into the optimization problem as this can have a significant impact on the uncertainty reducing decision. The results also indicated that the use of a RO or a SO model can lead to significantly different uncertainty reduction decisions.

CHAPTER 4

Connected Uncertainty

4.1. Introduction

Current optimization methods typically model adversarial uncertainties to be independent across periods. In many practical settings, however, the realized uncertainty in a period can affect subsequent uncertainty realizations. Such connections have been addressed previously in the context of unit commitment [72] and inventory control problems [101]. The goal of this chapter is to provide a general framework to directly model connected uncertainties, in particular when they are based on time series. To this end, we develop this modeling framework for both robust and distributionally robust optimization.

In many applications, the uncertainty at a time period depends on the realization of the uncertain parameters in previous periods. A prominent example is when autocorrelation amongst uncertainties is observed in decision-making settings. In queuing systems, e.g., Livny et al. [71] study the impact of correlations in inter-arrival or service times. Similarly, Balvers and Mitchell [8] study the optimal portfolio choice problem under the assumption of autocorrelated stock returns. In newsvendor problems, Alwan et al. [4] provide models with autoregressive demand for commodities. These problems leverage the connection amongst uncertain parameters and would benefit from the RO paradigm to address worst-case outcomes. This is important because the connection amongst uncertainties may amplify the worst-case outcomes. However, to the best of our knowledge, connected uncertainty models in the context of RO have not yet been studied in a general fashion.

We study the class of interdependent and adversarial uncertainties in RO problems while maintaining generality to ensure broad applicability. Specifically, we seek to provide a step towards modeling this family of problems with uncertainty sets that capture connections to previous realizations. We then extend this approach to the DRO perspective and provide reformulations for moment based ambiguity sets which depend on past realizations. To develop intuition, we begin with an example.

4.1.1. Example

Consider a stylized knapsack problem:

$$\begin{aligned} \max_{\mathbf{x}_t \in \mathcal{X}} \ & \sum_{t=1}^{T} \mathbf{c}_t^\top \mathbf{x}_t \\ \text{s.t.} \ & \sum_{t=1}^{T} \boldsymbol{\xi}_t^\top \mathbf{x}_t \leq B \quad \forall \boldsymbol{\xi}_t \in \mathcal{U}_t \quad \forall t = 1, \ldots, T, \end{aligned}$$

where $\mathbf{x}_t \in \mathcal{X}_t \subseteq \mathbb{R}^n \ \forall t = 1, \ldots, T$ are decision variables, $\mathbf{c}_t \in \mathbb{R}^n$ are known, $\boldsymbol{\xi}_t \in \mathbb{R}^n$ are uncertain coefficients, and B is the right-hand side (RHS) coefficient. In this setting, the uncertain $\boldsymbol{\xi}_t$ can vary in each period while being correlated with the past. Figure 4.1(a) illustrates some of the possible uncertainty realizations. In the paradigm of RO, these uncertainties can be modeled by uncertainty sets $\mathcal{U}_t$ in different ways. For clarity, consider sets $\mathcal{U}_t$ that are parametrized by their centers $\boldsymbol{\mu}_t$ and sizes r_t. In the following, we discuss three distinct models for $T = 3$.

- Often uncertainty sets across periods are modeled to be invariant as $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$ and $r_1 = r_2 = r_3$, resulting in $\mathcal{U}_1 = \mathcal{U}_2 = \mathcal{U}_3$ as illustrated in Figure 4.1(b). While this fixed model offers simplicity and computational advantages, it fails to capture all possible scenarios, when uncertainties are actually auto-correlated.

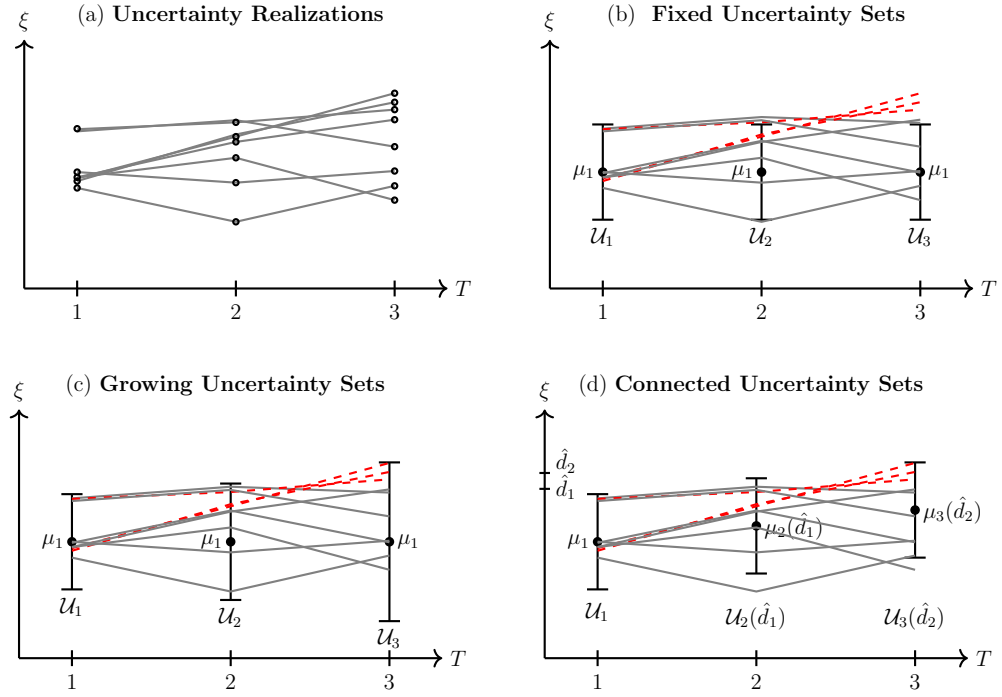


Figure 4.1. Uncertainties over time (a) can be modeled with (b) fixed, (c) growing, and (d) connected sets with μ_t updated using specific realizations $\hat{\xi}_t$.

- To overcome this limitation, $\mathcal{U}_t$ can be modeled to grow in size, i.e., $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$ and $r_1 \leq r_2 \leq r_3$, resulting in $\mathcal{U}_1 \subseteq \mathcal{U}_2 \subseteq \mathcal{U}_3$. Such growing uncertainty sets are sketched in Figure 4.1(c). The expansion, however, may render the solutions over conservative.
- In both previous models, the uncertainty set in each period is independent of the realizations in previous periods. We propose a new uncertainty set model $\mathcal{U}_t = \mathcal{U}_t(\boldsymbol{\xi}_{t-1})$ for $t > 1$, where the center of the set is correlated with the previous period uncertainty realization. Such a setting occurs, for example, when $\boldsymbol{\mu}_2(\boldsymbol{\xi}_1) = \mathbf{A}\boldsymbol{\xi}_1$ and $\boldsymbol{\mu}_3(\boldsymbol{\xi}_2) = \mathbf{A}\boldsymbol{\xi}_2$ with $\mathbf{A} = \alpha\mathbf{I}$, $\alpha > 0$, and $r_1 = r_2 = r_3$. Such sets are illustrated in Figure 4.1(d) for positively correlated centers. This model has the advantage of capturing autocorrelated uncertainties without increasing set size.

We will revisit this example in Section 4.4 with more detailed numerical analysis. For clarity, we now formally define the proposed uncertainty sets.

Definition 4.1.1. A connected uncertainty set *(CU set) has parameters that can depend on the realizations of the uncertainty in previous periods.*

Consequently, we call realizations from CU sets as *connected uncertainties*. We will provide specific examples of such sets for a variety of set structures.

4.1.2. Contributions

We study optimization problems over multiple time periods with both static and adaptive decisions where the uncertainty at each period is influenced by the uncertainty realizations in the past. Specifically, the contributions of this chapter are:

- We introduce the concept of *connected uncertainty sets* that offers generalizable insights. Specifically, we discuss uncertainties that depend linearly or quadratically on previous periods.
- We provide tractable reformulations with here-and-know decisions for common families of uncertainty models, namely i) polyhedral CU sets, where the right-hand side parameters of the set depend on previous periods, ii) ellipsoidal CU sets, where the center depends on the past, or (iii) ellipsoidal sets, where the covariance matrix depends on previous periods.
- We also provide tractable reformulations for affinely adaptive decisions for polyhedral and ellipsoidal uncertainty sets, which are commonly used sets in practical settings.
- We probe the proposed concept through numerical experiments on a stylized knapsack problem as it is applicable to a wide range of managerial and decision-making applications.

To clarify, note that such connected sets $\mathcal{U}_t(\boldsymbol{\xi}_{t-1})$ can also be expressed as a single joint uncertainty set over $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T$ to ease the reformulation. However, a key advantage of the proposed CU sets is their explicit dependence on previous realizations. This modeling power can be exploited, e.g., when the uncertainty stems from a time series. Furthermore, if an application requires the repeated solving of the problem, our approach prescribes how to update the sets over time.

The closest work to ours is application related. Lorca and Sun [72, 73] study multistage economic dispatch and unit commitment under uncertain renewable power generation. They incorporate a time series into uncertainty sets and solve the problem by cut generation, since large-scale unit commitment problems are known to be intractable. Our work is on general constraints, leverages duality theory for tractable reformulations, and can be used in a variety of applications.

4.2. Connected Uncertainty with RO

The concept of interstage dependence in multistage problems is well studied in stochastic optimization as it is naturally embedded in scenario trees and can be used to model complex state dependent processes such as financial volatility [54]. Some models assume independence, allowing decomposition algorithms and sharing of cuts within the same stage, e.g., L-shaped method [56]. In general, sharing cuts is not permitted when the scenarios are dependent unless the dependency follows simple time series models in which case these cuts can be modified and shared [34].

In RO, uncertainty models typically employ a single set for all periods. The popular budgeted uncertainty set models use one budget across periods to couple the uncertainty realizations [9, 57, 81]. Building upon this concept, the notion of dynamic uncertainty sets allows a time series model to capture fluctuating demand. The set at each period depends on realizations in

all previous periods [72, 73]. Similarly, in inventory management, Mamani et al. [76] studied correlated and non-identically distributed demand over multiple periods, and Ang et al. [6] investigated storage assignment.

In this chapter, we consider the dependence on past realizations for general purpose uncertainty sets rather than application specific sets. We reformulate a standard linear constraint over a sequence of uncertainty sets whose parameters (centers, covariances etc.) are connected across time.

4.2.1. Model

In a variety of applications such as cutting stock and packing problems, capital budgeting, and project selection etc., linear constraints are typically subject to connected uncertainties [41]. Since this effect can be confined to constraint coefficients, the goal of this section is to reformulate the constraint

$$\text{(C-RO)} \qquad \sum_{t=1}^{T} \boldsymbol{\xi}_t^\top \mathbf{x}_t \leq B \quad \forall \boldsymbol{\xi}_t \in \mathcal{U}_t(\boldsymbol{\xi}_{t-1}) \;\; \forall \boldsymbol{\xi}_1 \in \mathcal{U}_1 \;\; \forall t = 2, \ldots, T.$$

Here, $\boldsymbol{\xi}_t$ is the vector of uncertain coefficients, $\mathbf{x}_t$ denotes the decision variables, and B is a constant upper bound. In each period, $\boldsymbol{\xi}_t$ resides in a set $\mathcal{U}_t(\cdot)$, which may depend on $\boldsymbol{\xi}_{t-1}$. The robust counterpart of (C-RO) becomes

$$\text{(RC)} \qquad \max_{\boldsymbol{\xi}_1 \in \mathcal{U}_1} \{\boldsymbol{\xi}_1^\top \mathbf{x}_1 + \max_{\boldsymbol{\xi}_2 \in \mathcal{U}_2(\boldsymbol{\xi}_1)} \{\boldsymbol{\xi}_2^\top \mathbf{x}_2 + \cdots + \max_{\boldsymbol{\xi}_T \in \mathcal{U}_T(\boldsymbol{\xi}_{T-1})} \boldsymbol{\xi}_T^\top \mathbf{x}_T\}\} \leq B.$$

Each period t is affected by the worst-case realization of the uncertain parameter in period $t-1$.

In this chapter, we focus on CU sets, whose parameters linearly depend on the previous realization. Such uncertainties are prevalent in modeling popular applications, e.g., autoregressive models and minimum mean square error predictors (linear or jointly normal) [52]. In the

RO literature, when confidence regions of the uncertainty are available, ellipsoidal sets are commonly used, whereas when only bounds on the uncertainty are known, polyhedral uncertainty sets are used. Therefore, we discuss CU sets of ellipsoidal and polyhedral structures to extend the benefits of RO to multi-period settings. For ellipsoidal sets, we study the dependence of the set center and of the covariance matrix on the past, which may arise from either a Bayesian or frequentist update, or from a time series model. For polyhedral sets, we focus on the RHS set coefficients that may depend on previous realizations. This occurs when the magnitude or the location of the uncertainty is affected by past realizations.

4.2.2. Center Dependence of Ellipsoidal Sets

When uncertainties are Gaussian they can be modeled with ellipsoidal sets as Gaussian distributions have ellipsoidal contours. Confidence or predictive regions can also be naturally described by ellipsoids [29]. For such sets, there are three parameters that can be affected by previous uncertainty realizations: radius r, center $\boldsymbol{\mu}_t$, or covariance matrix $\boldsymbol{\Sigma}_t$. The case of radius dependence leads to a nested norm structure, resulting in nonconvex problems that are beyond our scope. Here, we discuss the setting where $\boldsymbol{\mu}_t$ depends on the previous period realizations, while r and $\boldsymbol{\Sigma}_t$ are constant. Such dependence arises, when, e.g., the demand is autoregressive [4]. Section 4.2.3 discusses $\boldsymbol{\Sigma}_t$ dependence while r and $\boldsymbol{\mu}_t$ are invariant.

When the set centers $\boldsymbol{\mu}_t$ are connected, the uncertainty set for each period is

$$\text{(E)} \qquad \mathcal{U}_t(\boldsymbol{\xi}_{t-1}) = \{\boldsymbol{\xi}_t \mid \boldsymbol{\xi}_t = \boldsymbol{\mu}_t(\boldsymbol{\xi}_{t-1}) + \mathbf{L}_t\mathbf{u}_t : \|\mathbf{u}_t\|_2 \leq r_t\},$$

where $\mathbf{L}_t\mathbf{L}_t^\top = \boldsymbol{\Sigma}_t$. If $\boldsymbol{\mu}_t$ is autoregressive and depends on the previous period realization as

$$\boldsymbol{\mu}_t(\boldsymbol{\xi}_{t-1}) = \mathbf{A}_t\boldsymbol{\mu}_{t-1}(\boldsymbol{\xi}_{t-2}) + \mathbf{F}_t\boldsymbol{\xi}_{t-1} + \mathbf{c}_t, \tag{4.1}$$

the constraint (C-RO) can be reformulated as following.

The nature of the connectedness motivates $\mathbf{A}_t$, $\mathbf{F}_t$ and $\mathbf{c}_t$, e.g., by autoregressive or Bayesian models, etc. We reformulate the uncertainty in a period assuming all previous realizations are known. The robust counterpart at t has to take future uncertainties into account. As such, all decisions are affected by succeeding decisions. We define recursive variables for each period $\tau \in \{1, \ldots, T-1\}$ as:

$$\begin{aligned}
\mathbf{y}_{T-\tau} &= \mathbf{x}_{T-\tau} + (\mathbf{F}_{T-\tau+1} + \mathbf{A}_{T-\tau+1})^\top \mathbf{y}_{T-\tau+1}, \text{ with } \mathbf{y}_T = \mathbf{x}_T, \\
C_{T-\tau} &= \mathbf{c}_{T-\tau+1}^\top \mathbf{y}_{T-\tau+1} + C_{T-\tau+1}, \text{ with } C_{T+1} = 0 \text{ and} \\
R_{T-\tau} &= r_{T-\tau} \| L_{T-\tau}^\top (\mathbf{x}_{T-\tau} + \mathbf{F}_{T-\tau+1} \mathbf{y}_{T-\tau+1}) \|_2 + R_{T-\tau+1} \text{ with } R_T = r_T \| \mathbf{L}_T^\top \mathbf{x}_T \|_2.
\end{aligned}$$

The aggregates $C_{T-\tau}$ and $R_{T-\tau}$ are functions of the variable $\mathbf{y}_{T-\tau}$ and represent the effect of the constants $\mathbf{c}_t$ in the update (4.1) and the 2-norm protection term, respectively.

Theorem 4.2.1. *The robust counterpart of constraint* (C-RO) *for the ellipsoidal set* (E) *is*

$$\boldsymbol{\mu}_1^\top \mathbf{y}_1 + C_2 + R_1 \leq B.$$

Proof. We start by reformulating (C-RO) for an arbitrary τ and then extend it to $\tau = T$. Let $s_{T-\tau} = \sum_{t=1}^{T-\tau} \boldsymbol{\xi}_t^\top \mathbf{x}_t$, then the robust counterpart of (C-RO) for τ is

$$s_{T-\tau} + \boldsymbol{\mu}_{T-\tau+1} (\boldsymbol{\xi}_{T-\tau})^\top \mathbf{y}_{T-\tau+1} + C_{T-\tau+2} + R_{T-\tau+1} \leq B. \tag{4.2}$$

We prove the statement for $\boldsymbol{\xi}_T$, then assume it to be true for $\boldsymbol{\xi}_{T-k+1}$, before proving it for $\boldsymbol{\xi}_{T-k}$.

Base case ($\tau = 1$). The constraint (C-RO) can be expanded as $s_{T-1} + \boldsymbol{\xi}_T^\top \mathbf{x}_T + C_{T+1} \leq B$. Because the constraint must hold for all $\mathbf{u}_T \in \{\mathbf{u}_T \mid \|\mathbf{u}_T\|_2 \leq r_T\}$, it also holds for the robust counterpart

$$s_{T-1} + \boldsymbol{\mu}_T (\boldsymbol{\xi}_{T-1})^\top \mathbf{x}_T + r_T \| \mathbf{L}_T^\top \mathbf{x}_T \|_2 + C_{T+1} \leq B,$$

as $\boldsymbol{\xi}_T = \boldsymbol{\mu}_T(\boldsymbol{\xi}_{T-1}) + \mathbf{L}_T\mathbf{u}_T$. With $\mathbf{y}_T = \mathbf{x}_T$, $C_{T+1} = 0$, and $R_T = r_T\|\mathbf{L}_T^\top \mathbf{x}_T\|_2$, we obtain the result

$$s_{T-1} + \boldsymbol{\mu}_T(\boldsymbol{\xi}_{T-1})^\top \mathbf{y}_T + C_{T+1} + R_T \leq B.$$

Inductive case ($\tau = k$). Assume that the reformulation (4.2) holds for $T - k + 1$. Then the robust counterpart of (C-RO) with respect to $\boldsymbol{\xi}_T, \boldsymbol{\xi}_{T-1}, \ldots, \boldsymbol{\xi}_{T-k+1}$ is

$$s_{T-k} + \boldsymbol{\mu}_{T-k+1}(\boldsymbol{\xi}_{T-k})^\top \mathbf{y}_{T-k+1} + C_{T-k+2} + R_{T-k+1} \leq B.$$

Substituting the mean and rearranging the terms, this can be expressed as

$$\begin{aligned} s_{T-k-1} &+ \boldsymbol{\xi}_{T-k}^\top(\mathbf{x}_{T-k} + \mathbf{F}_{T-k+1}^\top \mathbf{y}_{T-k+1}) + \boldsymbol{\mu}_{T-k}(\boldsymbol{\xi}_{T-k-1})^\top \mathbf{A}_{T-k+1}^\top \mathbf{y}_{T-k+1} \\ &+ \mathbf{c}_{T-k+1}^\top \mathbf{y}_{T-k+1} + C_{T-k+2} + R_{T-k+1} \leq B. \end{aligned}$$

Using the uncertainty set (E), this can be rewritten as

$$\begin{aligned} s_{T-k-1} &+ \boldsymbol{\mu}_{T-k}(\boldsymbol{\xi}_{T-k-1})^\top(\mathbf{x}_{T-k} + \mathbf{F}_{T-k+1}^\top \mathbf{y}_{T-k+1}) \\ &+ \boldsymbol{\mu}_{T-k}(\boldsymbol{\xi}_{T-k-1})^\top \mathbf{A}_{T-k+1}^\top \mathbf{y}_{T-k+1} + \mathbf{u}_{T-k}^\top \mathbf{L}_{T-k}^\top(\mathbf{x}_{T-k} + \mathbf{F}_{T-k+1}^\top \mathbf{y}_{T-k+1}) \\ &+ \mathbf{c}_{T-k+1}^\top \mathbf{y}_{T-k+1} + C_{T-k+2} + R_{T-k+1} \leq B. \end{aligned}$$

Taking the robust counterpart with respect to $\mathbf{u}_{T-k}$, we can write

$$\begin{aligned} s_{T-k-1} &+ \boldsymbol{\mu}_{T-k}(\boldsymbol{\xi}_{T-k-1})^\top(\mathbf{x}_{T-k} + \mathbf{F}_{T-k+1}^\top \mathbf{y}_{T-k+1} + \mathbf{A}_{T-k+1}^\top \mathbf{y}_{T-k+1}) \\ &+ r_{T-k}\|\mathbf{L}_{T-k}^\top(\mathbf{x}_{T-k} + \mathbf{F}_{T-k+1}^\top \mathbf{y}_{T-k+1})\|_2 \\ &+ \mathbf{c}_{T-k+1}^\top \mathbf{y}_{T-k+1} + C_{T-k+2} + R_{T-k+1} \leq B. \end{aligned}$$

With $\mathbf{y}_{T-k} = \mathbf{x}_{T-k} + \mathbf{F}_{T-k+1}^{\top}\mathbf{y}_{T-k+1} + \mathbf{A}_{T-k+1}^{\top}\mathbf{y}_{T-k+1}$ and the definitions of R_{T-k} and C_{T-k}, we obtain the desired result

$$s_{T-k-1} + \boldsymbol{\mu}_{T-k}(\boldsymbol{\xi}_{T-k-1})^{\top}\mathbf{y}_{T-k} + C_{T-k+1} + R_{T-k} \leq B.$$

This concludes the induction, and the final reformulation is obtained by substituting $\tau = T$. $\square$

In summary, when modeling uncertainty with connected centers, a recursive variable $\mathbf{y}_{T-\tau+1}$ is required to protect against the accumulating effect of future uncertainties. Current RO models, which model uncertainties at each period to be independent, neglect this effect. Thus they do not capture all uncertainties or lead to over conservative solutions. In order to optimally account for uncertainty over all periods, Theorem 4.2.1 prescribes how to use this $\mathbf{y}_{T-\tau+1}$ to modify the 2-norm in order to protect against uncertainty at each period. We will numerically demonstrate this result for connected centers on a knapsack application in Section 4.4 and show that they can improve constraint satisfaction and increase objective function value for any given level of constraint satisfaction, when compared to non-connected sets.

Remark: Alternatively, a joint uncertainty set can be constructed by the union of the individual uncertainty sets (E) over all time periods. This joint set is second-order cone representable as such the constraint (C-RO) can be tractably reformulated over this set as well.

4.2.3. Matrix Dependence of Ellipsoidal Sets

The previous section focused on changes in the location of ellipsoidal uncertainty sets. Many problems require models that allow for the shape of the set to vary, e.g., when the volatility in an uncertain process depends on past realizations. Such uncertainties are usually described by autoregressive conditional heteroskedastic models, e.g. in asset prices [27] or demand and sales growth of firms [68]. To this end, we model the covariance $\boldsymbol{\Sigma}_t$ of the ellipsoidal uncertainty

set to depend on previous realizations, while the radius r and mean $\boldsymbol{\mu}_t$ are predetermined. A general form of this dependence is

$$\boldsymbol{\Sigma}_{T-t+1}(\boldsymbol{\xi}_{T-t}) = a_t \boldsymbol{\Sigma}_{T-t}(\boldsymbol{\xi}_{T-t-1}) + f_t(\boldsymbol{\xi}_{T-t} - \boldsymbol{\mu}_{T-t})(\boldsymbol{\xi}_{T-t} - \boldsymbol{\mu}_{T-t})^\top + \mathbf{C}_t \ \ \forall t, \tag{4.3}$$

where $a_t \geq 0$, $f_t \geq 0$, $\mathbf{C}_t \succeq 0$, and $\boldsymbol{\Sigma}_1 \succeq 0$ are constants. This dependence captures frequentist and Bayesian paradigms, where the former assumes no prior distribution and the latter assumes one. Because the second term in (4.3) is rank 1 and $\mathbf{C}_t$ is positive semi-definite, the update (4.3) preserves the positive semi-definiteness of $\boldsymbol{\Sigma}_t$.

The quadratic dependence in (4.3) leads to nonlinear terms in the robust counterpart, obscuring an analytic reformulation. For this, Theorem 4.2.2 provides a conservative reformulation. Since the quadratic term can arise from a positive or a negative deviation from $\boldsymbol{\mu}_{T-t}$, we introduce a sign variable $\mathbf{n}^\top = (\mathbf{n}_1^\top, \mathbf{n}_2^\top, \ldots, \mathbf{n}_{T-1}^\top)$ with $\mathbf{n}_t = (n_{t,1}, n_{t,2}, \ldots, n_{t,t})$ to differentiate between these cases. Each $n_{t,k}$ can be 1 or -1, and $\mathbf{n}_{[T-\tau]}^\top = (\mathbf{n}_{T-\tau}^\top, \ldots, \mathbf{n}_{T-1}^\top)$. The set $\mathcal{N}$ consists of all possible $2^{\frac{T}{2}(T-1)}$ vectors $\mathbf{n}$, and $\mathcal{N}_{[T-\tau]}$ consists of all combinations of $\mathbf{n}_{[T-\tau]}$. For $\tau \in \{1, \ldots, T-1\}$, we also define

$$\mathbf{y}_{T-\tau}(\mathbf{n}_{[T-\tau]}) = \mathbf{x}_{T-\tau} + \sum_{t=1}^{\tau} n_{T-\tau,t} \cdot \mathbf{y}_{T-t+1}(\mathbf{n}_{[T-t+1]})\, r_{T-t+1} \sqrt{A_{\tau,t} f_\tau},$$

$$\text{with } \mathbf{y}_{T+1} = 0 \text{ and } \mathbf{y}_T = \mathbf{x}_T,$$

$$R_{T-\tau+1} = R_{T-\tau+2} + \sum_{t=1}^{\tau} r_{T-t+1} \sqrt{A_{\tau,t}\, \mathbf{y}_{T-t+1}(\mathbf{n}_{T-t+1})^\top \mathbf{C}_\tau \mathbf{y}_{T-t+1}(\mathbf{n}_{T-t+1})},$$

$$\text{with } R_{T+1} = 0, \text{and}$$

$$A_{\tau,t} = \prod_{j=t}^{\tau-1} a_j, \ t = 1, \ldots, \tau - 1,$$

$$\text{with } A_{\tau,\tau} = 1.$$

In what follows, we suppress the dependence of $\mathbf{y}_{T-\tau}$ on $\mathbf{n}_{[T-\tau]}$ and $\boldsymbol{\Sigma}_{t+1}$ on $\boldsymbol{\xi}_t$ for brevity. To reformulate (C-RO), let $\boldsymbol{\xi}_{T-\tau}$ for some $\tau \in \{1, \ldots, T-1\}$ reside in a set (E) with radius $r_{T-\tau}$, center $\boldsymbol{\mu}_{T-\tau}$, and covariance matrix $\boldsymbol{\Sigma}_{T-\tau}$ with $\boldsymbol{\Sigma}_{T-\tau} = \mathbf{L}_{T-\tau}\mathbf{L}_{T-\tau}^\top$ updated as in (4.3).

Theorem 4.2.2. *A conservative robust reformulation of* (C-RO) *is*

$$\sum_{t=1}^{T} \boldsymbol{\mu}_t^\top \mathbf{x}_t + \sum_{t=1}^{T} r_{T-t+1}\sqrt{A_{T,t}}\, \|\mathbf{L}_1^\top \mathbf{y}_{T-t+1}\|_2 + R_2 \leq B \;\; \forall \mathbf{n} \in \mathcal{N}.$$

Proof. The reformulation proceeds through induction. The robust counterpart of (C-RO) with respect to $\boldsymbol{\xi}_{T-\tau+1}, \ldots, \boldsymbol{\xi}_T$ is

$$s_{T-\tau} + \Theta_{T-\tau+1} + \sum_{t=1}^{\tau} r_{T-t+1}\sqrt{A_{\tau,t}\mathbf{y}_{T-t+1}^\top \boldsymbol{\Sigma}_{T-\tau+1}\mathbf{y}_{T-t+1}} + R_{T-\tau+2} \leq B,$$

$$\forall \mathbf{n}_{T-\tau+1} \in \mathcal{N}_{[T-\tau+1]},$$

$$\text{with } s_{T-\tau} = \sum_{t=1}^{T-\tau} \boldsymbol{\xi}_t^\top \mathbf{x}_t, \;\; \Theta_{T-\tau+1} = \sum_{t=T-\tau+1}^{T} \boldsymbol{\mu}_t^\top \mathbf{x}_t \text{ and } s_0 = 0.$$

The proof proceeds parallel to that of Theorem 4.2.1.

Base Case ($\tau = 1$). The constraint (C-RO) can be written as $s_{T-1} + \boldsymbol{\xi}_T^\top \mathbf{x}_T \leq B$. Substituting $\boldsymbol{\xi}_T = \boldsymbol{\mu}_T + \mathbf{L}_T\mathbf{u}_T$. (C-RO) must hold for all $\mathbf{u}_T$ with $\|\mathbf{u}_T\|_2 \leq r_T$, and thus it also holds for $s_{T-1} + \boldsymbol{\mu}_T^\top \mathbf{x}_T + r_T\|\mathbf{L}_T^\top \mathbf{x}_T\|_2 \leq B$. Using the equality $\|\mathbf{L}_T^\top \mathbf{x}_T\|_2 = \sqrt{\mathbf{x}_T^\top \boldsymbol{\Sigma}_T \mathbf{x}_T}$ and $\mathbf{y}_T = \mathbf{x}_T$, we obtain

$$s_{T-1} + \Theta_T + r_T\|\mathbf{L}_T^\top \mathbf{y}_T\|_2 \leq B \Leftrightarrow s_{T-1} + \Theta_T + r_T\sqrt{\mathbf{y}_T^\top \boldsymbol{\Sigma}_T \mathbf{y}_T} \leq B.$$

Since R_{T+1} is assumed to be zero, we have achieved the desired result.

Inductive case ($\tau = k$). Assume that the result is true for $T-k+1$, i.e., the reformulation with respect to $\boldsymbol{\xi}_T, \boldsymbol{\xi}_{T-1}, \dots \boldsymbol{\xi}_{T-k+1}$ is given by

$$s_{T-k} + \Theta_{T-k+1} + \sum_{t=1}^{k} r_{T-t+1}\sqrt{A_{k,t}\mathbf{y}_{T-t+1}^{\top}\boldsymbol{\Sigma}_{T-k+1}\mathbf{y}_{T-t+1}} + R_{T-k+2} \leq B$$
$$\forall \mathbf{n}_{T-k+1} \in \mathcal{N}_{[T-k+1]}.$$

Now, we have to prove that the reformulation with respect to $\boldsymbol{\xi}_{T-k}$ with $\tau = k+1$ is given by

$$s_{T-k-1} + \Theta_{T-k} + \sum_{t=1}^{k+1} r_{T-t+1}\sqrt{A_{k+1,t}\,\mathbf{y}_{T-t+1}^{\top}\boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-t+1}} + R_{T-k+1} \leq B$$
$$\forall \mathbf{n}_{T-k} \in \mathcal{N}_{[T-k]}.$$

Substituting $s_{T-k} = s_{T-k-1} + \boldsymbol{\xi}_{T-k}^{\top}\mathbf{x}_{T-k}$ and $\boldsymbol{\xi}_{T-k} = \boldsymbol{\mu}_{T-k} + \mathbf{L}_{T-k}\mathbf{u}_{T-k}$ in the inductive assumption

$$s_{T-k-1} + \Theta_{T-k+1} + \boldsymbol{\mu}_{T-k}^{\top}\mathbf{x}_{T-k} + \mathbf{u}_{T-k}^{\top}\mathbf{L}_{T-k}^{\top}\mathbf{x}_{T-k}$$
$$+ \sum_{t=1}^{k} r_{T-t+1}\sqrt{A_{k,t}\mathbf{y}_{T-t+1}^{\top}\boldsymbol{\Sigma}_{T-k+1}\mathbf{y}_{T-t+1}} + R_{T-k+2} \leq B \;\; \forall \mathbf{n}_{T-k+1} \in \mathcal{N}_{[T-k+1]}.$$

Utilizing (4.3) for $t = k$ and $\Theta_{T-k} = \boldsymbol{\mu}_{T-k}^{\top}\mathbf{x}_{T-k} + \Theta_{T-k+1}$, the constraint can be reformulated as

$$s_{T-k-1} + \Theta_{T-k} + \mathbf{u}_{T-k}^{\top}\mathbf{L}_{T-k}^{\top}\mathbf{x}_{T-k}$$
$$+ \sum_{t=1}^{k} r_{T-t+1}\Bigg(A_{k,t}a_k\mathbf{y}_{T-t+1}^{\top}\boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-t+1} + A_{k,t}f_k((\boldsymbol{\xi}_{T-k} - \boldsymbol{\mu}_{T-k})^{\top}\mathbf{y}_{T-t+1})^2$$
$$+ A_{k,t}\mathbf{y}_{T-t+1}^{\top}\mathbf{C}_k\mathbf{y}_{T-t+1}\Bigg)^{\frac{1}{2}} + R_{T-k+2} \leq B \;\; \forall \mathbf{n}_{T-k+1} \in \mathcal{N}_{[T-k+1]}.$$

Using the Cauchy-Schwarz inequality, a more conservative constraint is

$$s_{T-k-1} + \Theta_{T-k} + \mathbf{u}_{T-k}^{\top}\mathbf{L}_{T-k}^{\top}\mathbf{x}_{T-k} + \sum_{t=1}^{k} r_{T-t+1}\sqrt{A_{k,t}a_k\mathbf{y}_{T-t+1}^{\top}\boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-t+1}}$$
$$+ \sum_{t=1}^{k} r_{T-t+1}\sqrt{A_{k,t}f_k}\left|(\boldsymbol{\xi}_{T-k} - \boldsymbol{\mu}_{T-k})^{\top}\mathbf{y}_{T-t+1}\right|$$
$$+ \sum_{t=1}^{k} r_{T-t+1}\sqrt{A_{k,t}\mathbf{y}_{T-t+1}^{\top}\mathbf{C}_k\mathbf{y}_{T-t+1}} + R_{T-k+2} \leq B \;\; \forall \mathbf{n}_{T-k+1} \in \mathcal{N}_{[T-k+1]}.$$

The above constraint holds for $\left|(\boldsymbol{\xi}_{T-k} - \boldsymbol{\mu}_{T-k})^{\top}\mathbf{y}_{T-t+1}\right|$, if it holds for both the positive and the negative possible values. We use $n_{T-k,t}$, which can take values 1 or -1 to express the above as

$$s_{T-k-1} + \Theta_{T-k} + \mathbf{u}_{T-k}^{\top}\mathbf{L}_{T-k}^{\top}\mathbf{x}_{T-k} + \sum_{t=1}^{k} r_{T-t+1}\sqrt{A_{k,t}a_k\mathbf{y}_{T-t+1}^{\top}\boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-t+1}}$$
$$+ \sum_{t=1}^{k} n_{T-k,t} \cdot r_{T-t+1}\sqrt{A_{k,t}f_k}(\boldsymbol{\xi}_{T-k} - \boldsymbol{\mu}_{T-k})^{\top}\mathbf{y}_{T-t+1}$$
$$+ \sum_{t=1}^{k} r_{T-t+1}\sqrt{A_{k,t}\mathbf{y}_{T-t+1}^{\top}\mathbf{C}_k\mathbf{y}_{T-t+1}} + R_{T-k+2} \leq B$$
$$\forall n_{T-k,t} \in \{1,-1\},\ \forall t = 1,\ldots,k,\ \forall \mathbf{n}_{T-k+1} \in \mathcal{N}_{[T-k+1]}.$$

Note that the use of $n_{T-k,t}$ allows to write the 2^k constraints in this concise form. Combining the set $\mathcal{N}_{[T-k+1]}$ with $\{1,-1\}^k$, using $\boldsymbol{\xi}_{T-k} - \boldsymbol{\mu}_{T-k} = \mathbf{L}_{T-k}\mathbf{u}_{T-k}$, and collecting the terms involving

$\mathbf{u}_{T-k}^\top \mathbf{L}_{T-k}^\top$, we obtain

$$
\begin{aligned}
& s_{T-k-1} + \Theta_{T-k} + \mathbf{u}_{T-k}^\top \mathbf{L}_{T-k}^\top (\mathbf{x}_{T-k} + \sum_{t=1}^{k} n_{T-k,t} \cdot r_{T-t+1} \sqrt{A_{k,t} f_k} \mathbf{y}_{T-t+1}) \\
& + \sum_{t=1}^{k} r_{T-t+1} \sqrt{A_{k,t} a_k \mathbf{y}_{T-t+1}^\top \mathbf{\Sigma}_{T-k} \mathbf{y}_{T-t+1}} + \sum_{t=1}^{k} r_{T-t+1} \sqrt{A_{k,t} \mathbf{y}_{T-t+1}^\top \mathbf{C}_k \mathbf{y}_{T-t+1}} \\
& + R_{T-k+2} \leq B \ \ \forall \mathbf{n}_{T-k} \in \mathcal{N}_{[T-k]}.
\end{aligned}
$$

With the definition of $\mathbf{y}_{T-k}$, we can write

$$
\begin{aligned}
s_{T-k-1} + \Theta_{T-k} \quad + \quad & \mathbf{u}_{T-k}^\top \mathbf{L}_{T-k}^\top \mathbf{y}_{T-k} + \sum_{t=1}^{k} r_{T-t+1} \sqrt{A_{k,t} a_k \mathbf{y}_{T-t+1}^\top \mathbf{\Sigma}_{T-k} \mathbf{y}_{T-t+1}} \\
+ \quad & \sum_{t=1}^{k} r_{T-t+1} \sqrt{A_{k,t} \mathbf{y}_{T-t+1}^\top \mathbf{C}_k \mathbf{y}_{T-t+1}} + R_{T-k+2} \leq B \ \ \forall \mathbf{n}_{T-k} \in \mathcal{N}_{[T-k]}.
\end{aligned}
$$

Since this constraint must hold for all $\mathbf{u}_{T-k}$ with $\|\mathbf{u}_{T-k}\|_2 \leq r_{T-k}$, taking the maximum over $\mathbf{u}_{T-k}$ and substituting $\|\mathbf{L}_{T-k}^\top \mathbf{y}_{T-k}\|_2 = \sqrt{\mathbf{y}_{T-k}^\top \mathbf{\Sigma}_{T-k} \mathbf{y}_{T-k}}$, we obtain

$$
\begin{aligned}
& s_{T-k-1} + \Theta_{T-k} + r_{T-k} \sqrt{\mathbf{y}_{T-k}^\top \mathbf{\Sigma}_{T-k} \mathbf{y}_{T-k}} \\
& + \sum_{t=1}^{k} r_{T-t+1} \sqrt{A_{k,t} a_k \mathbf{y}_{T-t+1}^\top \mathbf{\Sigma}_{T-k} \mathbf{y}_{T-t+1}} \\
& + \sum_{t=1}^{k} r_{T-t+1} \sqrt{A_{k,t} \mathbf{y}_{T-t+1}^\top \mathbf{C}_k \mathbf{y}_{T-t+1}} + R_{T-k+2} \leq B \ \ \forall \mathbf{n}_{T-k} \in \mathcal{N}_{[T-k]}.
\end{aligned} \tag{4.4}
$$

Using $A_{k,t} a_k = \prod_{j=t}^{k-1} a_j a_k = A_{k+1,t}$, $A_{k+1,k+1} = 1$, and

$$
\begin{aligned}
& r_{T-k} \sqrt{A_{k+1,k+1} \mathbf{y}_{T-k}^\top \mathbf{\Sigma}_{T-k} \mathbf{y}_{T-k}} \\
& \qquad = r_{T-(k+1)+1} \sqrt{A_{k+1,k+1} \mathbf{y}_{T-(k+1)+1}^\top \mathbf{\Sigma}_{T-(k+1)+1} \mathbf{y}_{T-(k+1)+1}},
\end{aligned}
$$

we can rewrite (4.4) as

$$s_{T-k-1} + \Theta_{T-k} + \sum_{t=1}^{k+1} r_{T-t+1}\sqrt{A_{k+1,t}\ \mathbf{y}_{T-t+1}^{\top}\boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-t+1}} + \sum_{t=1}^{k} r_{T-t+1}\sqrt{A_{k,t}\mathbf{y}_{T-t+1}^{\top}\mathbf{C}_k\mathbf{y}_{T-t+1}} + R_{T-k+2} \leq B \ \ \forall \mathbf{n}_{T-k} \in \mathcal{N}_{[T-k]},$$

by extending the first summation. Using the definition of R_{T-k+1}, we obtain the desired result,

$$s_{T-k-1} + \Theta_{T-k} + \sum_{t=1}^{k+1} r_{T-t+1}\sqrt{A_{k+1,t}\ \mathbf{y}_{T-t+1}^{\top}\boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-t+1}} + R_{T-k+1} \leq B \quad \forall \mathbf{n}_{T-k} \in \mathcal{N}_{[T-k]}.$$

The complete reformulation follows immediately by substituting $\tau = T$. □

The reformulation in Theorem 4.2.2 consists of a sum of 2-norms, making it computationally tractable, and hence attractive for large-scale problems with few periods. Similar to center dependence in Section 4.2.2, $\mathbf{y}_{T-\tau}$ plays the key role in adapting the protection against connected matrices. This advances the conventional RO paradigm because the presence of connectedness allows current worst-case realizations to affect future uncertainties, and Theorem 4.2.2 describes how $\mathbf{y}_{T-\tau}$ modifies the 2-norm through the summation $\sum_{t=1}^{\tau} n_{T-\tau,t} \cdot \mathbf{y}_{T-t+1}(\mathbf{n}_{[T-t+1]})\ r_{T-t+1}\sqrt{f_\tau A_{\tau,t}}$ in order to incorporate this affect.

Summary for Ellipsoidal Sets. For center and matrix dependence, we showed how $\boldsymbol{\mu}_t(\boldsymbol{\xi}_{t-1})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\xi}_{t-1})$ can be affected by past uncertainties, and that the worst case is driven by future uncertainties through $\mathbf{y}_{T-t}$, as in (4.3) and (4.4), respectively. These results explicitly reveal the effect of connected uncertainties, namely that the past dependence establishes the *location* (for $\boldsymbol{\mu}_t$ dependence via (4.1) or for $\boldsymbol{\Sigma}_t$ dependence via (4.3)), while the future connections determine the *direction* of the worst case (via theorems 4.2.1 for $\boldsymbol{\mu}_t$ or 4.2.2 for $\boldsymbol{\Sigma}_t$).

4.2.4. Polyhedral Sets

When the upper and lower bounds of the uncertainty are known, box sets are commonly used, e.g., in inventory management under demand uncertainty [89]. To reduce the inherent conservatism, a budget can be placed on the number of uncertain components [18]. Such polyhedral sets provide robustness while maintaining tractability. Here, we extend these models to the CU setting. Specifically, we consider multi-period robust problems, where the uncertain constraint coefficients of each period reside in a polyhedral CU set. We model the RHS set parameters to depend on the realization of the previous period uncertainty. Such settings may arise, e.g., when the bounds on the uncertain demand in an inventory problem depend on past demand realizations [see 76].

To reformulate the constraint (C-RO), let the uncertain $\boldsymbol{\xi}_t$ reside in a polyhedral set

$$\text{(P)} \qquad \mathcal{U}_t(\boldsymbol{\xi}_{t-1}) = \{\boldsymbol{\xi}_t \mid \mathbf{G}_t\boldsymbol{\xi}_t \geq \mathbf{g}_t + \boldsymbol{\Delta}_t\boldsymbol{\xi}_{t-1}\},$$

where the matrix $\boldsymbol{\Delta}_t$ is application based, e.g., from time series models. Here, without loss of generality, the parameters of $\mathcal{U}_1$ are considered as known ($\boldsymbol{\Delta}_1 = \mathbf{0}$). This setting resembles the introductory Example 3, and the following theorem reformulates (C-RO).

Theorem 4.2.3. *The robust counterpart of constraint* (C-RO) *under the uncertainty set* (P) *is*

$$\begin{aligned}
&\sum_{t=1}^{T} \mathbf{q}_t^\top \mathbf{g}_t \leq B \\
&\mathbf{q}_t^\top \mathbf{G}_t = \mathbf{x}_t^\top + \mathbf{q}_{t+1}^\top \boldsymbol{\Delta}_{t+1} && \forall t = 1, \ldots, T, \\
&\mathbf{q}_t \leq 0 && \forall t = 1, \ldots, T,
\end{aligned}$$

where $\boldsymbol{\Delta}_1 = 0$ *and* $\boldsymbol{\Delta}_{T+1} = 0$.

The proof leverages duality and is relegated to Appendix A.1. In this theorem, the term $\mathbf{q}_t^\top \mathbf{g}_t$ safeguards against all uncertainty realizations. The dual variable $\mathbf{q}_t$ is appropriately adjusted by $\mathbf{q}_t^\top \mathbf{G}_t = (\mathbf{x}_t + \mathbf{\Delta}_{t+1}^\top \mathbf{q}_{t+1})^\top$ to account for uncertainties. The connectedness, i.e., the nonzero $\mathbf{\Delta}_{t+1}$, contributes the second term $\mathbf{\Delta}_{t+1}^\top \mathbf{q}_{t+1}$. It is worthwhile pointing out that if the uncertainty dependence occurs on the left-hand side (LHS) (e.g., $\widetilde{\mathcal{U}}_t(\boldsymbol{\xi}_{t-1}) = \{\boldsymbol{\xi}_t \mid \mathbf{G}_t(\boldsymbol{\xi}_{t-1})\boldsymbol{\xi}_t \geq \mathbf{g}_t\}$), then the corresponding reformulation requires dual variables that are functions of $\boldsymbol{\xi}_{t-1}$, resulting in an infinite dimensional optimization problem.

In conclusion, theorems 4.2.1, 4.2.2 and 4.2.3 highlight that connected uncertainties can be modeled in a natural fashion via CU sets, which also enable tractability. These results show that modifying the decision variable in order to account for future worst cases is an instrumental component of reformulating robust optimization problems with CU sets.

4.2.5. Extensions to Affinely adaptive Decisions

In previous sections, we focused on the effect of connected uncertainties on *static* and here-and-now decisions. In many problems, decisions can also adapt to previously revealed uncertainties. Adjustable robust optimization (ARO) problems were introduced by Ben-Tal et al. [11] and have applications in inventory management [89] and unit commitment [73], amongst others. ARO problems, however, are known to be NP-complete, which can be circumvented by leveraging affine or piece-wise static decision rules [51].

In this section, we extend the notion of modeling with CU sets to *adaptive* decisions and provide reformulations for a two-period ARO problem under affine decision rules. Such a problem

can be expressed as

$$
\text{(CU-ARO)} \quad
\begin{aligned}
\min_{\substack{\mathbf{x}_1 \\ \mathbf{x}_2(\boldsymbol{\xi}_1)}} \max_{\substack{\boldsymbol{\xi}_1 \in \mathcal{U}_1 \\ \boldsymbol{\xi}_2 \in \mathcal{U}_2(\boldsymbol{\xi}_1)}} \quad & \mathbf{c}_1^\top \mathbf{x}_1 + \mathbf{c}_2^\top \mathbf{x}_2(\boldsymbol{\xi}_1) & & \\
\text{s.t.} \quad & \mathbf{A}_{11}\mathbf{x}_1 \geq \mathbf{b}_1(\boldsymbol{\xi}_1) & & \forall \boldsymbol{\xi}_1 \in \mathcal{U}_1 \\
& \mathbf{A}_{21}\mathbf{x}_1 + \mathbf{A}_{22}\mathbf{x}_2(\boldsymbol{\xi}_1) \geq \mathbf{b}_2(\boldsymbol{\xi}_2) & & \forall \boldsymbol{\xi}_2 \in \mathcal{U}_2(\boldsymbol{\xi}_1)\, \forall \boldsymbol{\xi}_1 \in \mathcal{U}_1 \\
& \mathbf{x}_2(\boldsymbol{\xi}_1) \geq \mathbf{0} & & \forall \boldsymbol{\xi}_1 \in \mathcal{U}_1 \\
& \mathbf{x}_1 \geq \mathbf{0}. & &
\end{aligned}
$$

We focus on polyhedral CU sets, where the RHS depends on the previous period realization as

$$\mathcal{U}_1 = \{\boldsymbol{\xi}_1 \mid \mathbf{G}_1\boldsymbol{\xi}_1 \geq \mathbf{g}_1\}, \qquad \mathcal{U}_2(\boldsymbol{\xi}_1) = \{\boldsymbol{\xi}_2 \mid \mathbf{G}_2\boldsymbol{\xi}_2 \geq \mathbf{g}_2 + \boldsymbol{\Delta}_1\boldsymbol{\xi}_1\}.$$

For this setting, the following theorem shows the impact of CU sets with affine decision policies on the fully adaptive decisions $\mathbf{x}_2(\boldsymbol{\xi}_1) = \mathbf{x}_2^0 + \mathbf{X}_2\boldsymbol{\xi}_1$. For brevity, we incorporated the constant term $\mathbf{x}_2^0$ into the matrix $\mathbf{X}_2$ by modifying the uncertainty set and setting the first component of $\boldsymbol{\xi}_1$ to be equal to 1. Then we can write $\mathbf{x}_2(\boldsymbol{\xi}_1) = \mathbf{X}_2\boldsymbol{\xi}_1$.

Theorem 4.2.4. *The two-period adjustable RO problem* (CU-ARO) *has a tractable reformulation, when the uncertainty affects the RHS linearly and the fully adaptive decisions are replaced by affine decision rules.*

Proof. We replace $\mathbf{x}_2(\boldsymbol{\xi}_1)$ with the affine decision rule $\mathbf{x}_2(\boldsymbol{\xi}_1) = \mathbf{X}_2\boldsymbol{\xi}_1$ and expand $\mathbf{b}_1(\boldsymbol{\xi}_1) = \mathbf{B}_1\boldsymbol{\xi}_1$ and $\mathbf{b}_2(\boldsymbol{\xi}_2) = \mathbf{B}_2\boldsymbol{\xi}_2$. We focus on reformulating the second constraint which is affected by

the connected uncertainty and whose robust problem is

$$\begin{aligned}
&\max_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2} \mathbf{B}_{2,i}^\top \boldsymbol{\xi}_2 - [\mathbf{A}_{22}\mathbf{X}_2]_i^\top \boldsymbol{\xi}_1 \\
&\text{s.t.} \quad \mathbf{G}_1\boldsymbol{\xi}_1 \geq \mathbf{g}_1 \\
&\qquad \mathbf{G}_2\boldsymbol{\xi}_2 \geq \mathbf{g}_2 + \Delta_1\boldsymbol{\xi}_1.
\end{aligned}$$

Using the dual, the complete second constraint of (CU-ARO) is given by

$$\begin{aligned}
\mathbf{A}_{21}\mathbf{x}_1 &\geq \mathbf{P}^\top\mathbf{g}_1 + \mathbf{Q}^\top\mathbf{g}_2 \\
\mathbf{P}^\top\mathbf{G}_1 + \mathbf{Q}^\top\boldsymbol{\Delta}_1 &= \mathbf{B}_2 \\
\mathbf{Q}^\top\mathbf{G}_2 &= -\mathbf{A}_{22}\mathbf{X}_2 \\
\mathbf{P}, \mathbf{Q} &\leq \mathbf{0}.
\end{aligned}$$

The columns of $\mathbf{P}$ and $\mathbf{Q}$ correspond to dual variables of the original problem. The remaining constraints in (CU-ARO) can be reformulated similarly, leading to a tractable reformulation. □

We can similarly obtain tractable reformulations for ellipsoidal CU sets via second order conic programs, with the center of the ellipsoid being a linear function of the first period uncertainty realization. The details of this reformulation are provided in Appendix A.2.

Theorem 4.2.4 provides a tractable reformulation for connected RHS uncertainty, covering a broad range of applications. This extends the result of modeling with standard sets for adaptive setting, where such reformulations are possible for RHS uncertainty, to CU sets. When uncertainties affect the LHS, even problems with standard and non-connected sets are difficult to reformulate [12]. Therefore, we forego their extension to the CU setting.

When uncertainties follow unknown distributions, the concept of connected sets can also be applied to multi-period distributionally robust problems. Moment-based ambiguity sets lend themselves naturally, where, e.g., the mean or covariance depends on uncertainty realizations

from the previous period. In a parallel fashion to the RO setting, we now discuss modeling with CU sets in a distributional environment for static solutions.

4.3. Connected Uncertainty with DRO

In distributionally robust settings, the uncertainty set models all distributions that satisfy the set specifying constraints. For CU sets, the parameters of these constraints may depend on previous realizations. Consider the following example of CU sets that are specified by distributional moments,

$$\widetilde{\mathcal{U}}_1 = \Big\{ P_1 \in \mathcal{M} \Big| P_1(\boldsymbol{\xi}_1 \in \Xi_1) = 1,\ \big|\mathbb{E}_{P_1}[\boldsymbol{\xi}_1] - \boldsymbol{\mu}_1\big| \leq \boldsymbol{\delta}_1, \\ \mathbb{E}_{P_1}[(\boldsymbol{\xi}_1 - \boldsymbol{\mu}_1)(\boldsymbol{\xi}_1 - \boldsymbol{\mu}_1)^\top] \preceq \boldsymbol{\Sigma}_1 \Big\},$$

$$\widetilde{\mathcal{U}}_2(\boldsymbol{\xi}_1) = \Big\{ P_{2|1} \in \mathcal{M} \Big| P_{2|1}(\boldsymbol{\xi}_2 \in \Xi_2) = 1,\ \big|\mathbb{E}_{P_{2|1}}[\boldsymbol{\xi}_2] - \boldsymbol{\mu}_2(\boldsymbol{\xi}_1)\big| \leq \boldsymbol{\delta}_2, \\ \mathbb{E}_{P_{2|1}}[(\boldsymbol{\xi}_2 - \boldsymbol{\mu}_2^0)(\boldsymbol{\xi}_2 - \boldsymbol{\mu}_2^0)^\top] \preceq \boldsymbol{\Sigma}_2 \Big\},$$

where $P_{2|1}$ describes the conditional distribution of $\boldsymbol{\xi}_2$ given the realization of $\boldsymbol{\xi}_1$. In these uncertainty sets, the conditional mean $\mathbb{E}_{P_{2|1}}[\boldsymbol{\xi}_2]$ is bounded by $\boldsymbol{\mu}_2(\boldsymbol{\xi}_1) + \boldsymbol{\delta}_2$ and $\boldsymbol{\mu}_2(\boldsymbol{\xi}_1) - \boldsymbol{\delta}_2$ and the covariance matrix by $\boldsymbol{\Sigma}_2$, with the bounds updated based on the previous realization. These sets naturally describe settings in which the uncertainty is modeled using time series. In what follows, we consider distributional uncertainty sets to be *connected*, when the moments at any given period depend on the realizations from previous periods, as presented in the example of $\widetilde{\mathcal{U}}_1$ and $\widetilde{\mathcal{U}}_2(\boldsymbol{\xi}_1)$.

The aim of this section is to reformulate the constraint

$$\text{(C-DRO)} \qquad \mathbb{E}_P[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \boldsymbol{\xi}_t)] \leq B.$$

The expectation $\mathbb{E}_P[\cdot]$ is taken over the joint distribution P of all $\boldsymbol{\xi}_t$, and $h_t(\mathbf{x}_t, \boldsymbol{\xi}_t)$ is a function of the decision variable $\mathbf{x}_t \in \mathbb{R}^{n_t}$ and the uncertain parameter $\boldsymbol{\xi}_t \in \mathbb{R}^m$. Unless specified, we do not make any assumptions on the structure of $h_t(\cdot,\cdot)$ beyond regularity conditions required for the existence of integrals. The dimension of $\boldsymbol{\xi}_t$ shall be constant over time for the clarity of exposition. Each $\boldsymbol{\xi}_t$ has a distribution that lies in a different uncertainty set, which depends on the previous realization $\boldsymbol{\xi}_{t-1}$.

Remark: While DRO problems with connected uncertainty sets appear similar to robust Markov decision processes (MDPs), they are different in nature. In robust MDPs, the timing of decision-making and uncertainty realization alternates, whereas in DRO with CU sets all decisions (over the entire time horizon) are made at the very first period, anticipating future uncertainties and before any realization of the uncertainty occurs. This is necessary because in many applications, longterm decisions have to be taken without full knowledge of future uncertainties, e.g., in risk management settings.

In what follows, we first provide a general reformulation for the constraint (C-DRO) before showing that the tractability can be improved with a conservative approximation.

4.3.1. Mean dependence

To reformulate (C-DRO) for moment based CU sets, consider the sets in which the bounds on the first moment parameter depend on the previous realization as $\boldsymbol{\mu}_t(\boldsymbol{\xi}_{t-1}) = \mathbf{A}_t\boldsymbol{\xi}_{t-1} + \mathbf{b}_t$, given by

$$\text{(D)} \quad \widetilde{\mathcal{U}}_t(\boldsymbol{\xi}_{t-1}) = \Big\{ P_{t|t-1} \in \mathcal{M} \Big| P_{t|t-1}(\boldsymbol{\xi}_t \in \Xi_t) = 1,\ \big|\mathbb{E}_{P_{t|t-1}}[\boldsymbol{\xi}_t] - \boldsymbol{\mu}_t(\boldsymbol{\xi}_{t-1})\big| \leq \boldsymbol{\delta}_t, \\ \mathbb{E}_{P_{t|t-1}}[(\boldsymbol{\xi}_t - \boldsymbol{\mu}_t^0)(\boldsymbol{\xi}_t - \boldsymbol{\mu}_t^0)^\top] \preceq \boldsymbol{\Sigma}_t \Big\}.$$

This set contains the distributions of $\boldsymbol{\xi}_t$ conditioned on $\boldsymbol{\xi}_{t-1}$. The parameter $\boldsymbol{\mu}_t^0$ denotes a constant estimate of the mean and is different than the true mean $\mathbb{E}_{P_{t|t-1}}[\boldsymbol{\xi}_t]$, which is unknown and bounded by the set. To prevent the dependence of the second moment terms on the previous realization, we also assume $\boldsymbol{\mu}_t^0$ to be different from $\boldsymbol{\mu}_t(\boldsymbol{\xi}_{t-1})$. A possible value for $\boldsymbol{\mu}_t^0$ can be the unconditional mean of a time series at time t. The robust counterpart of constraint (C-DRO) can be expressed with the following proposition

Proposition 4.3.1. *Given the sets $\widetilde{\mathcal{U}}_1, \ldots, \widetilde{\mathcal{U}}_T(\boldsymbol{\xi}_{T-1})$ and their joint uncertainty set $\widetilde{\mathcal{U}}$, we have the following. Constraint* (C-DRO)*, given by*

$$\sup_{P \in \widetilde{\mathcal{U}}} \mathbb{E}_P \left[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \boldsymbol{\xi}_t) \right] \leq B$$

is equivalent to

$$\begin{aligned} \sup_{P_1 \in \widetilde{\mathcal{U}}_1} \mathbb{E}_{P_1} \Bigg[h_1(\mathbf{x}_1, \boldsymbol{\xi}_1) + \sup_{P_{2|1} \in \widetilde{\mathcal{U}}_2(\boldsymbol{\xi}_1)} \Bigg\{ \mathbb{E}_{P_{2|1}} \Bigg[h_2(\mathbf{x}_2, \boldsymbol{\xi}_2) + \ldots \\ + \sup_{P_{T|T-1} \in \widetilde{\mathcal{U}}_T(\boldsymbol{\xi}_{T-1})} \Big\{ \mathbb{E}_{P_{T|T-1}} \Big[h_T(\mathbf{x}_T, \boldsymbol{\xi}_T) \Big] \Big\} \Bigg] \Bigg\} \Bigg] \leq B. \end{aligned} \tag{4.5}$$

The proof of Proposition 4.3.1 is provided in Appendix A.3. To ease the exposition, consider the function

$$g_t(\mathbf{x}_{[t+1:T]}, \boldsymbol{\xi}_t) \coloneqq \sup_{P_{t+1|t} \in \widetilde{\mathcal{U}}_{t+1}(\boldsymbol{\xi}_t)} \mathbb{E}_{P_{t+1|t}}[h_{t+1}(\mathbf{x}_{t+1}, \boldsymbol{\xi}_{t+1}) + g_{t+1}(\mathbf{x}_{[t+2:T]}, \boldsymbol{\xi}_{t+1})]. \tag{4.6}$$

For brevity, we denote $\mathbf{M}_t \equiv \boldsymbol{\Sigma}_t - \boldsymbol{\mu}_t^0 (\boldsymbol{\mu}_t^0)^\top$, and use the compact notations $\widetilde{p}_t, \widetilde{\mathbf{q}}_t^u, \widetilde{\mathbf{q}}_t^l$ and $\widetilde{\mathbf{R}}_t$ to denote variables $p_t(\boldsymbol{\xi}_{t-1}), \mathbf{q}_t^u(\boldsymbol{\xi}_{t-1}), \mathbf{q}_t^l(\boldsymbol{\xi}_{t-1})$ and $\mathbf{R}_t(\boldsymbol{\xi}_{t-1})$, which are functions of the previous period realization to ensure compactness. The following theorem provides the reformulation.

Theorem 4.3.1. *The constraint* (C-DRO) *can be reformulated as*

$$
\begin{aligned}
p_1 + (\mathbf{q}_1^u - \mathbf{q}_1^l)^\top \boldsymbol{\mu}_1 + (\mathbf{q}_1^u + \mathbf{q}_1^l)^\top \boldsymbol{\delta}_1 + \mathbf{R}_1 \cdot \mathbf{M}_1 &\leq B \\
\alpha_t(\boldsymbol{\xi}_{t-1}, \boldsymbol{\xi}_t) + \boldsymbol{\beta}_t(\boldsymbol{\xi}_{t-1}, \boldsymbol{\xi}_t)^\top \boldsymbol{\xi}_t + \boldsymbol{\xi}_t^\top \widetilde{\mathbf{R}}_t \boldsymbol{\xi}_t - h_t(\mathbf{x}_t, \boldsymbol{\xi}_t) &\geq 0 \\
&\forall (\boldsymbol{\xi}_{t-1}, \boldsymbol{\xi}_t) \in \Xi_{t-1} \times \Xi_t \ \ \forall t \\
\widetilde{p}_T + (\widetilde{\mathbf{q}}_T^u - \widetilde{\mathbf{q}}_T^l - 2\widetilde{\mathbf{R}}_T \boldsymbol{\mu}_T^0)^\top \boldsymbol{\xi}_T + \boldsymbol{\xi}_T^\top \widetilde{\mathbf{R}}_T \boldsymbol{\xi}_T - h_T(\mathbf{x}_T, \boldsymbol{\xi}_T) &\geq 0 \\
&\forall (\boldsymbol{\xi}_{T-1}, \boldsymbol{\xi}_T) \in \Xi_{T-1} \times \Xi_T \\
\widetilde{\mathbf{q}}_t^u, \widetilde{\mathbf{q}}_t^l \geq 0, \ \widetilde{\mathbf{R}}_t \succeq 0 \ \ \forall \boldsymbol{\xi}_{t-1} \in \Xi_{t-1} \\
\widetilde{\mathbf{q}}_T^u, \widetilde{\mathbf{q}}_T^l \geq 0, \ \widetilde{\mathbf{R}}_T \succeq 0 \ \ \forall \boldsymbol{\xi}_{T-1} \in \Xi_{T-1},
\end{aligned}
$$

where $t = 1, \ldots, T-1$ *and*

$$
\begin{aligned}
\alpha_i(\boldsymbol{\xi}_{i-1}, \boldsymbol{\xi}_i) &= \widetilde{p}_i - \widetilde{p}_{i+1} - (\widetilde{\mathbf{q}}_{i+1}^u - \widetilde{\mathbf{q}}_{i+1}^l)^\top \mathbf{b}_{i+1} - \widetilde{\mathbf{R}}_{i+1} \cdot \mathbf{M}_{i+1} - (\widetilde{\mathbf{q}}_{i+1}^u + \widetilde{\mathbf{q}}_{i+1}^l)^\top \boldsymbol{\delta}_{i+1} \\
\boldsymbol{\beta}_i(\boldsymbol{\xi}_{i-1}, \boldsymbol{\xi}_i) &= \widetilde{\mathbf{q}}_i^u - \widetilde{\mathbf{q}}_i^l - 2\widetilde{\mathbf{R}}_i \boldsymbol{\mu}_i^0 - \mathbf{A}_{i+1}^\top \widetilde{\mathbf{q}}_{i+1}^u + \mathbf{A}_{i+1}^\top \widetilde{\mathbf{q}}_{i+1}^l.
\end{aligned}
$$

Proof. The proof proceeds by induction. We first provide the reformulation for $t = 1$, then assume it to be true for $t = k$, before proving it for $t = k + 1$.

Base case ($t = 1$). The original constraint (C-DRO) can be expressed as

$$
\begin{aligned}
\sup_{P_1 \in \widetilde{\mathcal{U}}_1} \mathbb{E}_{P_1}[h_1(\mathbf{x}_1, \boldsymbol{\xi}_1) + \sup_{P_{2|1} \in \widetilde{\mathcal{U}}_2(\boldsymbol{\xi}_1)} \mathbb{E}_{P_{2|1}}[h_2(\mathbf{x}_2, \boldsymbol{\xi}_2) + \ldots \\
+ \sup_{P_{T|T-1} \in \widetilde{\mathcal{U}}_T(\boldsymbol{\xi}_{T-1})} \mathbb{E}[h_T(\mathbf{x}_T, \boldsymbol{\xi}_T)]]] \leq B,
\end{aligned}
$$

which can be compressed as $\sup_{P_1 \in \widetilde{\mathcal{U}}_1} \mathbb{E}_{P_1}[h_1(\mathbf{x}_1, \boldsymbol{\xi}_1) + g_1(\mathbf{x}_{[2:T]}, \boldsymbol{\xi}_1)] \leq B$. This optimization problem can be expressed as the following moment problem

$$\begin{aligned}
\sup_{P_1 \in \mathcal{M}} \; & \int_{\Xi_1} (h_1(\mathbf{x}_1, \boldsymbol{\xi}_1) + g_1(\mathbf{x}_{[2:T]}, \boldsymbol{\xi}_1)) dP_1 \\
& \int_{\Xi_1} dP_1 = 1 \\
& \boldsymbol{\mu}_1 - \boldsymbol{\delta}_1 \leq \int_{\Xi_1} \boldsymbol{\xi}_1 dP_1 \leq \boldsymbol{\mu}_1 + \boldsymbol{\delta}_1 \\
& \int_{\Xi_1} (\boldsymbol{\xi}_1 - \boldsymbol{\mu}_1^0)(\boldsymbol{\xi}_1 - \boldsymbol{\mu}_1^0)^\top dP_1 \preceq \boldsymbol{\Sigma_1}.
\end{aligned}$$

The last constraint is equivalent to $\mathbb{E}_{P_{t|t-1}}[\boldsymbol{\xi}_t \boldsymbol{\xi}_t^\top] - 2\boldsymbol{\mu}_t^0 \mathbb{E}_{P_{t|t-1}}[\boldsymbol{\xi}_t^\top] \preceq \boldsymbol{\Sigma}_t - \boldsymbol{\mu}_t^0 (\boldsymbol{\mu}_t^0)^\top$. Thus the dual problem is given by

$$\begin{aligned}
\inf_{p_1, \mathbf{q}_1^u, \mathbf{q}_1^l, \mathbf{R}_1} \; & p_1 + (\mathbf{q}_1^u)^\top (\boldsymbol{\mu}_1 + \boldsymbol{\delta}_1) - (\mathbf{q}_1^l)^\top (\boldsymbol{\mu}_1 - \boldsymbol{\delta}_1) + \mathbf{R}_1 \cdot \mathbf{M}_1 \\
\text{s.t.} \quad & p_1 + (\mathbf{q}_1^u - \mathbf{q}_1^l)^\top \boldsymbol{\xi}_1 - 2(\boldsymbol{\mu}_1^0)^\top \mathbf{R}_1 \boldsymbol{\xi}_1 + \boldsymbol{\xi}_1^\top \mathbf{R}_1 \boldsymbol{\xi}_1 \geq h_1(\mathbf{x}_1, \boldsymbol{\xi}_1) \\
& \qquad\qquad + g_1(\mathbf{x}_{[2:T]}, \boldsymbol{\xi}_1) \;\; \forall \boldsymbol{\xi}_1 \in \Xi_1 \\
& \mathbf{q}_1^u, \mathbf{q}_1^l \geq 0 \\
& \mathbf{R}_1 \succeq 0,
\end{aligned}$$

which proves the base case.

Inductive case ($t = k$). For the $t = k + 1$ reformulation, consider the function

$$g_k(\mathbf{x}_{[k+1:T]}, \boldsymbol{\xi}_k) = \sup_{P_{k+1|k} \in \widetilde{\mathcal{U}}_{k+1}(\boldsymbol{\xi}_k)} \mathbb{E}_{P_{k+1|k}}[h_{k+1}(\mathbf{x}_{k+1}, \boldsymbol{\xi}_{k+1}) + g_{k+1}(\mathbf{x}_{[k+2:T]}, \boldsymbol{\xi}_{k+1})]. \tag{4.7}$$

Now, the third constraint in the reformulation of Theorem 4.3.1 for $t = k$ can be expressed as

$$\widetilde{p}_k + (\widetilde{\mathbf{q}}_k^u - \widetilde{\mathbf{q}}_k^l - 2\widetilde{\mathbf{R}}_k\boldsymbol{\mu}_k^0)^\top\boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\top\widetilde{\mathbf{R}}_k\boldsymbol{\xi}_k \geq h_k(\mathbf{x}_k, \boldsymbol{\xi}_k)$$
$$+ \sup_{P_{k+1|k}\in\widetilde{\mathcal{U}}_{k+1}(\boldsymbol{\xi}_k)} \mathbb{E}_{P_{k+1|k}}[h_{k+1}(\mathbf{x}_{k+1}, \boldsymbol{\xi}_{k+1}) + g_{k+1}(\mathbf{x}_{[k+2:T]}, \boldsymbol{\xi}_{k+1})]$$
$$\forall\boldsymbol{\xi}_{k-1} \in \Xi_{k-1} \;\; \forall\boldsymbol{\xi}_k \in \Xi_k.$$

Using the dual of (4.7), we can write the above constraint as the following two constraints

$$\widetilde{p}_k + (\widetilde{\mathbf{q}}_k^u - \widetilde{\mathbf{q}}_k^l - 2\widetilde{\mathbf{R}}_k\boldsymbol{\mu}_k^0)^\top\boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\top\widetilde{\mathbf{R}}_k\boldsymbol{\xi}_k$$
$$\geq h_k(\mathbf{x}_k, \boldsymbol{\xi}_k) + \widetilde{p}_{k+1} + \widetilde{\mathbf{R}}_{k+1}\cdot\mathbf{M}_{k+1} + (\widetilde{\mathbf{q}}_{k+1}^u - \widetilde{\mathbf{q}}_{k+1}^l)^\top\boldsymbol{\mu}_{k+1}(\boldsymbol{\xi}_k)$$
$$+ (\widetilde{\mathbf{q}}_{k+1}^u + \widetilde{\mathbf{q}}_{k+1}^l)^\top\boldsymbol{\delta}_{k+1} \qquad \forall\boldsymbol{\xi}_{k-1} \in \Xi_{k-1} \;\; \forall\boldsymbol{\xi}_k \in \Xi_k$$
$$\widetilde{p}_{k+1} + (\widetilde{\mathbf{q}}_{k+1}^u - \widetilde{\mathbf{q}}_{k+1}^l - 2\widetilde{\mathbf{R}}_{k+1}\boldsymbol{\mu}_{k+1}^0)^\top\boldsymbol{\xi}_{k+1} + \boldsymbol{\xi}_{k+1}^\top\widetilde{\mathbf{R}}_{k+1}\boldsymbol{\xi}_{k+1}$$
$$\geq h_{k+1}(\mathbf{x}_{k+1}, \boldsymbol{\xi}_{k+1}) + g_{k+1}(\mathbf{x}_{[k+2:T]}, \boldsymbol{\xi}_{k+1}) \qquad \forall\boldsymbol{\xi}_k \in \Xi_k \;\; \forall\boldsymbol{\xi}_{k+1} \in \Xi_{k+1}$$
$$\widetilde{\mathbf{q}}_{k+1}^u, \widetilde{\mathbf{q}}_{k+1}^l \geq 0, \; \widetilde{\mathbf{R}}_{k+1} \succeq 0 \qquad \forall\boldsymbol{\xi}_k \in \Xi_k.$$

Substituting $\boldsymbol{\mu}_{k+1}(\boldsymbol{\xi}_k) = \mathbf{A}_{k+1}\boldsymbol{\xi}_k + \mathbf{b}_{k+1}$, we rearrange the first constraint as

$$p_k + (\widetilde{\mathbf{q}}_k^u - \widetilde{\mathbf{q}}_k^l - 2\widetilde{\mathbf{R}}_k\boldsymbol{\mu}_k^0 - \mathbf{A}_{k+1}^\top\widetilde{\mathbf{q}}_{k+1}^u + \mathbf{A}_{k+1}^\top\widetilde{\mathbf{q}}_{k+1}^l)^\top\boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\top\widetilde{\mathbf{R}}_k\boldsymbol{\xi}_k$$
$$\geq h_k(\mathbf{x}_k, \boldsymbol{\xi}_k) + \widetilde{p}_{k+1} + (\widetilde{\mathbf{q}}_{k+1}^u - \widetilde{\mathbf{q}}_{k+1}^l)^\top\mathbf{b}_{k+1} + (\widetilde{\mathbf{q}}_{k+1}^u + \widetilde{\mathbf{q}}_{k+1}^l)^\top\boldsymbol{\delta}_{k+1}$$
$$+ \widetilde{\mathbf{R}}_{k+1}\cdot\mathbf{M}_{k+1} \qquad \forall\boldsymbol{\xi}_{k-1} \in \Xi_{k-1} \;\; \forall\boldsymbol{\xi}_k \in \Xi_k,$$

which can be written in a more compact form as

$$\alpha_k(\boldsymbol{\xi}_{k-1},\boldsymbol{\xi}_k)+\boldsymbol{\beta}_k(\boldsymbol{\xi}_{k-1},\boldsymbol{\xi}_k)^\top\boldsymbol{\xi}_k+\boldsymbol{\xi}_k^\top\widetilde{\mathbf{R}}_k\boldsymbol{\xi}_k-h_k(\mathbf{x}_k,\boldsymbol{\xi}_k)\geq 0$$
$$\forall\boldsymbol{\xi}_{k-1}\in\Xi_{k-1}\ \ \forall\boldsymbol{\xi}_k\in\Xi_k.$$

We can now give the complete set of constraints for $t=k+1$ as

$$p_1+(\mathbf{q}_1^u-\mathbf{q}_1^l)^\top\boldsymbol{\mu}_1+(\mathbf{q}_1^u+\mathbf{q}_1^l)^\top\boldsymbol{\delta}_1+\mathbf{R}_1\cdot\mathbf{M}_1\leq B$$
$$\alpha_t(\boldsymbol{\xi}_{t-1},\boldsymbol{\xi}_t)+\boldsymbol{\beta}_t(\boldsymbol{\xi}_{t-1},\boldsymbol{\xi}_t)^\top\boldsymbol{\xi}_t+\boldsymbol{\xi}_t^\top\widetilde{\mathbf{R}}_t\boldsymbol{\xi}_t-h_t(\mathbf{x}_t,\boldsymbol{\xi}_t)\geq 0$$
$$\forall\boldsymbol{\xi}_{t-1}\in\Xi_{t-1}\ \forall\boldsymbol{\xi}_t\in\Xi_t,\ \ \forall t$$
$$\widetilde{p}_{k+1}+(\widetilde{\mathbf{q}}_{k+1}^u-\widetilde{\mathbf{q}}_{k+1}^l-2\widetilde{\mathbf{R}}_{k+1}\boldsymbol{\mu}_{k+1}^0)^\top\boldsymbol{\xi}_{k+1}+\boldsymbol{\xi}_{k+1}^\top\widetilde{\mathbf{R}}_{k+1}\boldsymbol{\xi}_{k+1}\geq h_{k+1}(\mathbf{x}_{k+1},\boldsymbol{\xi}_{k+1})$$
$$+g_{k+1}(\mathbf{x}_{[k+2:T]},\boldsymbol{\xi}_{k+1})\ \ \forall\boldsymbol{\xi}_k\in\Xi_k\quad\forall\boldsymbol{\xi}_{k+1}\in\Xi_{k+1}.$$
$$\widetilde{\mathbf{q}}_t^u,\widetilde{\mathbf{q}}_t^l\geq 0\quad\forall\boldsymbol{\xi}_{t-1}\in\Xi_{t-1}\ \forall t$$
$$\widetilde{\mathbf{R}}_t\succeq 0\quad\forall\boldsymbol{\xi}_{t-1}\in\Xi_{t-1}\ \forall t,$$

The complete reformulation of the constraint (4.5) can be obtained by applying the induction up to $t=T$, where

$$g_{T-1}(\mathbf{x}_T,\boldsymbol{\xi}_{T-1})=\sup_{P_{T|T-1}\in\widetilde{\mathcal{U}}_T(\boldsymbol{\xi}_{T-1})}\mathbb{E}_{P_{T|T-1}}h_T(\mathbf{x}_T,\boldsymbol{\xi}_T).$$

□

Theorem 4.3.1 provides a prescription for how to modify the protection term, when uncertainties are connected. Notice that the variables in Theorem 4.3.1 depend on the uncertainty realization in the previous period. Furthermore, the variables α_i and $\boldsymbol{\beta}_i$ conjoin the variables of the current period with that of the future period. Therefore, Theorem 4.3.1 modifies the

protection in order to account for the dependence amongst the uncertainties. That means when, for example, autocorrelated uncertainties are modeled in the traditional (non-connected) way, the protection terms are not appropriately modified, which can lead to constraint violations.

4.3.2. Conservative Reformulation

Since the reformulation in Theorem 4.3.1 is an infinite dimensional optimization problem, which can be computationally challenging, the following theorem provides a conservative reformulation that is tractable.

Theorem 4.3.2. *The constraint* (C-DRO) *can be conservatively reformulated as*

$$
\begin{aligned}
& p_1 + (\mathbf{q}_1^u - \mathbf{q}_1^l)^\top \boldsymbol{\mu}_1 + (\mathbf{q}_1^u + \mathbf{q}_1^l)^\top \boldsymbol{\delta}_1 + \mathbf{R}_1 \cdot \mathbf{M}_1 \leq B \\
& \qquad \alpha_t + \boldsymbol{\beta}_t^\top \boldsymbol{\xi}_t + \boldsymbol{\xi}_t^\top \mathbf{R}_t \boldsymbol{\xi}_t - h_t(\mathbf{x}_t, \boldsymbol{\xi}_t) \geq 0 \qquad \forall \boldsymbol{\xi}_t \in \Xi_t \;\; \forall t = 1, \ldots, T-1 \\
& p_T + (\mathbf{q}_T^u - \mathbf{q}_T^l - 2\mathbf{R}_T \boldsymbol{\mu}_T^0)^\top \boldsymbol{\xi}_T + \boldsymbol{\xi}_T^\top \mathbf{R}_T \boldsymbol{\xi}_T \geq h_T(\mathbf{x}_T, \boldsymbol{\xi}_T) \qquad \forall \boldsymbol{\xi}_T \in \Xi_T \\
& \qquad \mathbf{q}_t^u, \mathbf{q}_t^l \geq 0, \; \mathbf{R}_t \succeq 0 \qquad \forall t = 1, \ldots, T,
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha_i &= p_i - p_{i+1} - (\mathbf{q}_{i+1}^u - \mathbf{q}_{i+1}^l)^\top \mathbf{b}_{i+1} - \mathbf{R}_{i+1} \cdot \mathbf{M}_{i+1} - (\mathbf{q}_{i+1}^u + \mathbf{q}_{i+1}^l)^\top \boldsymbol{\delta}_{i+1} \\
\boldsymbol{\beta}_i &= \mathbf{q}_i^u - \mathbf{q}_i^l - 2\mathbf{R}_i \boldsymbol{\mu}_i^0 - \mathbf{A}_{i+1}^\top \mathbf{q}_{i+1}^u + \mathbf{A}_{i+1}^\top \mathbf{q}_{i+1}^l.
\end{aligned}
$$

Proof. The proof proceeds by induction.

Base case ($t = 1$). The original constraint (C-DRO) can be expressed as

$$\sup_{P_1 \in \widetilde{\mathcal{U}}_1} \mathbb{E}_{P_1}[h_1(\mathbf{x}_1, \boldsymbol{\xi}_1) + \sup_{P_{2|1} \in \widetilde{\mathcal{U}}_2(\boldsymbol{\xi}_1)} \mathbb{E}_{P_{2|1}}[h_2(\mathbf{x}_2, \boldsymbol{\xi}_2) + \ldots + \sup_{P_{T|T-1} \in \widetilde{\mathcal{U}}_T(\boldsymbol{\xi}_{T-1})} \mathbb{E}[h_T(\mathbf{x}_T, \boldsymbol{\xi}_T)]]] \leq B,$$

which can be shortened to $\sup_{P_1 \in \widetilde{\mathcal{U}}_1} \mathbb{E}_{P_1}[h_1(\mathbf{x}_1, \boldsymbol{\xi}_1) + g_1(\mathbf{x}_{[2:T]}, \boldsymbol{\xi}_1)] \leq B$, and rewritten as

$$\begin{aligned} \sup_{P_1 \in \mathcal{M}} & \int_{\Xi_1} \left(h_1(\mathbf{x}_1, \boldsymbol{\xi}_1) + g_1(\mathbf{x}_{[2:T]}, \boldsymbol{\xi}_1)\right) dP_1 \\ & \int_{\Xi_1} dP_1 = 1 \\ & \boldsymbol{\mu}_1 - \boldsymbol{\delta}_1 \leq \int_{\Xi_1} \boldsymbol{\xi}_1 dP_1 \leq \boldsymbol{\mu}_1 + \boldsymbol{\delta}_1 \\ & \int_{\Xi_1} (\boldsymbol{\xi}_1 - \boldsymbol{\mu}_1^0)(\boldsymbol{\xi}_1 - \boldsymbol{\mu}_1^0)^\top dP_1 \preceq \boldsymbol{\Sigma_1}. \end{aligned}$$

The dual of this moment problem proves the base case via

$$\begin{aligned} \inf_{p_1, \mathbf{q}_1^u, \mathbf{q}_1^l, \mathbf{R}_1} \quad & p_1 + (\mathbf{q}_1^u)^\top(\boldsymbol{\mu}_1 + \boldsymbol{\delta}_1) - (\mathbf{q}_1^l)^\top(\boldsymbol{\mu}_1 - \boldsymbol{\delta}_1) + \mathbf{R}_1 \cdot \mathbf{M}_1 \\ \text{s.t.} \quad & p_1 + (\mathbf{q}_1^u - \mathbf{q}_1^l)^\top \boldsymbol{\xi}_1 - 2(\boldsymbol{\mu}_1^0)^\top \mathbf{R}_1 \boldsymbol{\xi}_1 + \boldsymbol{\xi}_1^\top \mathbf{R}_1 \boldsymbol{\xi}_1 \geq h_1(\mathbf{x}_1, \boldsymbol{\xi}_1) + g_1(\mathbf{x}_{[2:T]}, \boldsymbol{\xi}_1) \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall \boldsymbol{\xi}_1 \in \Xi_1 \\ & \mathbf{q}_1^u, \mathbf{q}_1^l \geq 0, \ \mathbf{R}_1 \succeq 0. \end{aligned}$$

Inductive case ($t = k$). We assume the constraints in Theorem 4.3.2 holds for $t = k$ and prove the reformulation for $t = k + 1$. These constraints can be expressed as

$$
\begin{aligned}
& p_1 + (\mathbf{q}_1^u - \mathbf{q}_1^l)^\top \boldsymbol{\mu}_1 + (\mathbf{q}_1^u + \mathbf{q}_1^l)^\top \boldsymbol{\delta}_1 + \mathbf{R}_1 \cdot \mathbf{M}_1 \leq B \\
& \qquad \alpha_t + \boldsymbol{\beta}_t^\top \boldsymbol{\xi}_t + \boldsymbol{\xi}_t^\top \mathbf{R}_t \boldsymbol{\xi}_t - h_t(\mathbf{x}_t, \boldsymbol{\xi}_t) \geq 0 \qquad \forall \boldsymbol{\xi}_t \in \Xi_t, \ \ \forall t \\
& \quad p_k + (\mathbf{q}_k^u - \mathbf{q}_k^l - 2\mathbf{R}_k \boldsymbol{\mu}_k^0)^\top \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\top \mathbf{R}_k \boldsymbol{\xi}_k \geq h_k(\mathbf{x}_k, \boldsymbol{\xi}_k) + g_k(\mathbf{x}_{[k+1:T]}, \boldsymbol{\xi}_k) \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall \boldsymbol{\xi}_k \in \Xi_k,
\end{aligned}
$$

where $t = 1, \ldots, k-1$ and α_t and $\boldsymbol{\beta}_t$ are given in the theorem. For $t = k + 1$, consider the function

$$
g_k(\mathbf{x}_{[k+1:T]}, \boldsymbol{\xi}_k) = \sup_{P_{k+1|k} \in \widetilde{\mathcal{U}}_{k+1}(\boldsymbol{\xi}_k)} \mathbb{E}_{P_{k+1|k}}[h_{k+1}(\mathbf{x}_{k+1}, \boldsymbol{\xi}_{k+1}) + g_{k+1}(\mathbf{x}_{[k+2:T]}, \boldsymbol{\xi}_{k+1})],
$$

which is bounded above by its dual for any feasible $(p_{k+1}, \mathbf{q}_{k+1}^u, \mathbf{q}_{k+1}^l, \mathbf{R}_{k+1}) \in \mathcal{P}_{k+1}$ as

$$
\begin{aligned}
g_k(\mathbf{x}_{[k+1:T]}, \boldsymbol{\xi}_k) \leq\ & p_{k+1} + \mathbf{R}_{k+1} \cdot \mathbf{M}_{k+1} \\
& + (\mathbf{q}_{k+1}^u - \mathbf{q}_{k+1}^l)^\top \boldsymbol{\mu}_{k+1}(\boldsymbol{\xi}_k) + (\mathbf{q}_{k+1}^u + \mathbf{q}_{k+1}^l)^\top \boldsymbol{\delta}_{k+1} \qquad \forall \boldsymbol{\xi}_k \in \Xi_k.
\end{aligned}
$$

Then the last constraint in the reformulation for $t = k$ can be conservatively expressed as

$$
\begin{aligned}
& p_k + (\mathbf{q}_k^u - \mathbf{q}_k^l - 2\mathbf{R}_k \boldsymbol{\mu}_k^0)^\top \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\top \mathbf{R}_k \boldsymbol{\xi}_k \\
& \qquad \geq h_k(\mathbf{x}_k, \boldsymbol{\xi}_k) + p_{k+1} + \mathbf{R}_{k+1} \cdot \mathbf{M}_{k+1} + (\mathbf{q}_{k+1}^u - \mathbf{q}_{k+1}^l)^\top \boldsymbol{\mu}_{k+1}(\boldsymbol{\xi}_k) \\
& \qquad + (\mathbf{q}_{k+1}^u + \mathbf{q}_{k+1}^l)^\top \boldsymbol{\delta}_{k+1} \qquad\qquad\qquad\qquad \forall \boldsymbol{\xi}_k \in \Xi_k \qquad (4.8) \\
& p_{k+1} + (\mathbf{q}_{k+1}^u - \mathbf{q}_{k+1}^l - 2\mathbf{R}_{k+1} \boldsymbol{\mu}_{k+1}^0)^\top \boldsymbol{\xi}_{k+1} + \boldsymbol{\xi}_{k+1}^\top \mathbf{R}_{k+1} \boldsymbol{\xi}_{k+1} \\
& \qquad \geq h_{k+1}(\mathbf{x}_{k+1}, \boldsymbol{\xi}_{k+1}) + g_{k+1}(\mathbf{x}_{[k+2:T]}, \boldsymbol{\xi}_{k+1}) \qquad \forall \boldsymbol{\xi}_{k+1} \in \Xi_{k+1}.
\end{aligned}
$$

Using $\boldsymbol{\mu}_{k+1}(\boldsymbol{\xi}_k) = \mathbf{A}_{k+1}\boldsymbol{\xi}_k + \mathbf{b}_{k+1}$, we can rearrange the first constraint as

$$\begin{aligned} p_k + (\mathbf{q}_k^u - \mathbf{q}_k^l - 2\mathbf{R}_k\boldsymbol{\mu}_k^0 - \mathbf{A}_{k+1}^\top\mathbf{q}_{k+1}^u + \mathbf{A}_{k+1}^\top\mathbf{q}_{k+1}^l)^\top\boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\top\mathbf{R}_k\boldsymbol{\xi}_k \\ \geq h_k(\mathbf{x}_k, \boldsymbol{\xi}_k) + p_{k+1} + (\mathbf{q}_{k+1}^u - \mathbf{q}_{k+1}^l)^\top\mathbf{b}_{k+1} + (\mathbf{q}_{k+1}^u + \mathbf{q}_{k+1}^l)^\top\boldsymbol{\delta}_{k+1} + \mathbf{R}_{k+1}\cdot\mathbf{M}_{k+1} \\ \forall\boldsymbol{\xi}_k \in \Xi_k, \end{aligned}$$

which can be written compactly as $\alpha_k + \boldsymbol{\beta}_k^\top\boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\top\mathbf{R}_k\boldsymbol{\xi}_k - h_k(\mathbf{x}_k, \boldsymbol{\xi}_k) \geq 0 \;\; \forall\boldsymbol{\xi}_k \in \Xi_k$. With this, the complete set of constraints for $t = k+1$ are

$$\begin{aligned} p_1 + (\mathbf{q}_1^u - \mathbf{q}_1^l)^\top\boldsymbol{\mu}_1 + (\mathbf{q}_1^u + \mathbf{q}_1^l)^\top\boldsymbol{\delta}_1 + \mathbf{R}_1\cdot\mathbf{M}_1 \leq B \\ \alpha_t + \boldsymbol{\beta}_t^\top\boldsymbol{\xi}_t + \boldsymbol{\xi}_t^\top\mathbf{R}_t\boldsymbol{\xi}_t - h_t(\mathbf{x}_t, \boldsymbol{\xi}_t) \geq 0 \\ \forall\boldsymbol{\xi}_{t-1} \in \Xi_{t-1} \;\; \forall\boldsymbol{\xi}_t \in \Xi_t, \;\; \forall t \\ p_{k+1} + (\mathbf{q}_{k+1}^u - \mathbf{q}_{k+1}^l - 2\mathbf{R}_{k+1}\boldsymbol{\mu}_{k+1}^0)^\top\boldsymbol{\xi}_{k+1} + \boldsymbol{\xi}_{k+1}^\top\mathbf{R}_{k+1}\boldsymbol{\xi}_{k+1} \geq h_{k+1}(\mathbf{x}_{k+1}, \boldsymbol{\xi}_{k+1}) \\ + g_{k+1}(\mathbf{x}_{[k+2:T]}, \boldsymbol{\xi}_{k+1}) \;\; \forall\boldsymbol{\xi}_{k+1} \in \Xi_{k+1}. \end{aligned}$$

The complete reformulation of the robust counterpart of (C-DRO) can be obtained by repeating the induction up to $t = T$, using

$$g_{T-1}(\mathbf{x}_T, \boldsymbol{\xi}_{T-1}) = \sup_{P_{T|T-1}\in\widetilde{\mathcal{U}}_T(\boldsymbol{\xi}_{T-1})} \mathbb{E}_{P_{T|T-1}} h_T(\mathbf{x}_T, \boldsymbol{\xi}_T).$$

□

In the reformulation of Theorem 4.3.2, α_i and $\boldsymbol{\beta}_i$ conjoin the variables of the current and the future periods, which is analogous to the results of theorems 4.2.1 and 4.3.1. This modification of the protection term accounts for the connection between the periods. When the support set is an ellipsoid and the objective function is piecewise linear concave in $\boldsymbol{\xi}_t$, it is possible to use

the S-Lemma to reformulate the constraints of the problems as semi-definite constraints, which can then be solved using commercially available solvers.

Remark: While Theorem 4.3.2 provides a conservative reformulation of the constraint (4.5), under certain conditions, it is also possible for this reformulation to be exact. This tightness occurs, when $g_{k-1}(\mathbf{x}_{[k:T]}, \boldsymbol{\xi}_{k-1})$ is a linear function of the uncertain component $\boldsymbol{\xi}_{k-1}$ over the set Ξ_{k-1}. The linearity arises when there exists a single optimal solution for all $\boldsymbol{\xi}_{k-1}$. Note that the verification of this condition requires solving a sequence of bilinear optimization problems, which is challenging.

In Section 4.5, we numerically demonstrate the advantages of CU sets through a portfolio optimization problem, where the realized wealth exhibits a lower standard deviation, supporting the notion of robust decisions. In the next two sections, we probe the performance of the proposed CU sets modeling frameworks through two stylized RO and DRO problems. Each demonstration serves to model a broad range of decision-making settings. For a clean experiment, randomly generated data is used to directly relate the findings to the properties of our proposed models.

4.4. RO Application: Knapsack Problem

Knapsack problems offer an insightful benchmark for optimization problems because they arise as sub-problems in many common applications [41, 77]. Consider a knapsack problem with known objective but uncertain constraint coefficients (a.k.a. weights). The problem spans two periods, each with an uncertain weight, where the second period weight value depends on the first realized weight. These uncertain parameters arise from an auto-regressive model. Such a

problem can be described by

$$\text{(KS)} \qquad \begin{aligned} \max_{\mathbf{x}_1,\mathbf{x}_2} \ & \mathbf{c}_1^\top \mathbf{x}_1 + \mathbf{c}_2^\top \mathbf{x}_2 \\ \text{s.t.} \ & \boldsymbol{\xi}_1^\top \mathbf{x}_1 + \boldsymbol{\xi}_2^\top \mathbf{x}_2 \leq B \qquad \forall \boldsymbol{\xi}_2 \in \mathcal{U}_2(\boldsymbol{\xi}_1), \ \forall \boldsymbol{\xi}_1 \in \mathcal{U}_1 \\ & \mathbf{x}_1 \in \{0,1\}^{m_1}, \ \mathbf{x}_2 \in \{0,1\}^{m_2}. \end{aligned}$$

The two binary decisions $\mathbf{x}_1$ and $\mathbf{x}_2$ with known objective $\mathbf{c}_1, \mathbf{c}_2$ and uncertain weight coefficients $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ correspond to periods one and two. Both decisions are taken before either of the weights are realized, i.e., $\mathbf{x}_1$ and $\mathbf{x}_2$ are here-and-now decisions. The uncertain parameters $\boldsymbol{\xi}_t$ for $t = 1, 2$ are modeled to reside in ellipsoidal sets with given covariance matrices and known first period center $\boldsymbol{\mu}_1$. The uncertainty dependence can be modeled by allowing the center of the second period ellipsoid $\boldsymbol{\mu}_2(\boldsymbol{\xi}_1)$ to depend on the realization of the first period weights $\boldsymbol{\xi}_1$ as

$$\boldsymbol{\mu}_2(\boldsymbol{\xi}_1) = \boldsymbol{\Phi}\boldsymbol{\mu}_1 + \boldsymbol{\Psi}\boldsymbol{\xi}_1.$$

Here, the parameters $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are matrices capturing the dependence and which may be determined through time series modeling. We emphasize that the solutions to problem (KS) are static, while the uncertainties are connected and as such stagewise. This uncertainty model is parallel to the discussion in Section 4.2.2.

Given the mean of either period $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2(\boldsymbol{\xi}_1)$, the residual uncertainties $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ in

$$\boldsymbol{\xi}_1 = \boldsymbol{\mu}_1 + \boldsymbol{\epsilon}_1 \text{ and } \ \boldsymbol{\xi}_2 = \boldsymbol{\Phi}\boldsymbol{\mu}_1 + \boldsymbol{\Psi}\boldsymbol{\xi}_1 + \boldsymbol{\epsilon}_2 \tag{4.9}$$

are characterized by a normal distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}$ such that $\mathbf{L}\mathbf{L}^\top = \boldsymbol{\Sigma}$. Then the corresponding *first period* uncertainty set is

$$\mathcal{U}_1 = \{\boldsymbol{\xi}_1 \mid \boldsymbol{\xi}_1 = \boldsymbol{\mu}_1 + \mathbf{L}\mathbf{u} \ : \ \|\mathbf{u}\|_2 \leq r_1\}.$$

The second period mean given the first period realization $\boldsymbol{\mu}_2(\boldsymbol{\xi}_1)$ has the same covariance matrix. Consequently, the *second period* CU set is

$$\mathcal{U}_2(\boldsymbol{\xi}_1) = \{\boldsymbol{\xi}_2 \mid \boldsymbol{\xi}_2 = \boldsymbol{\Phi}\boldsymbol{\mu}_1 + \boldsymbol{\Psi}\boldsymbol{\xi}_1 + \mathbf{L}\mathbf{w},\ \ \|\mathbf{w}\|_2 \leq r_2\}\,.$$

In the experiment we compare the performance of the connected model to the standard RO model, namely when uncertainty dependence is not taken into account. The first period has the same uncertainty set $\mathcal{U}_1$. For the second period set, the parameters are $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1$ and $\mathbf{L}\mathbf{L}^\top = \boldsymbol{\Sigma_2} = \boldsymbol{\Sigma}$. Therefore, the *non-connected* (NC) second period set is

$$\mathcal{U}_{2,\mathrm{NC}} = \{\boldsymbol{\xi}_2 \mid \boldsymbol{\xi}_2 = \boldsymbol{\mu}_2 + \mathbf{L}\mathbf{w},\ \ \|\mathbf{w}\|_2 \leq r_2\}\,.$$

Note that the key difference between $\mathcal{U}_2(\boldsymbol{\xi}_1)$ and $\mathcal{U}_{2,\mathrm{NC}}$ is that the center of $\mathcal{U}_2(\boldsymbol{\xi}_1)$ is $\boldsymbol{\Phi}\boldsymbol{\mu}_1 + \boldsymbol{\Psi}\boldsymbol{\xi}_1$ while that of $\mathcal{U}_{2,\mathrm{NC}}$ is $\boldsymbol{\mu}_2$, i.e., not updated according to the realization of $\boldsymbol{\xi}_1$. The covariance matrices for both remain the same. We now describe the experimental setting.

Computational Setup. We reformulate the robust problems as a mixed integer linear program (MILP) for the knapsack problem and as a semi definite program (SDP) for the portfolio optimization problem. These are then modeled using the JuMP library in the Julia programming language v0.6. The solvers used a Gurobi for the MILP and Mosek for the SDP. The experiments are run on a machine with an Intel Core i5 processor, with 8 GB RAM.

Numerical Experiments. Four experimental modules are conducted for uncertain $\boldsymbol{\xi}_1$, and $\boldsymbol{\xi}_2$, which are generated by (4.9) for fixed values of $\boldsymbol{\Phi} = \mathbf{I}$ and $\boldsymbol{\Psi} = \lambda\mathbf{I}$ with the identity matrix $\mathbf{I}$ and a scalar $\lambda = 0.5$. The parameter $\boldsymbol{\mu}_1$ is set to $\mathbf{e}$, and $\boldsymbol{\Sigma}$ is generated randomly. We consider a problem with 20 items in each period $m_1 = m_2 = 20$, following these steps.

(i) Generate k samples of $\mathbf{c}_1, \mathbf{c}_2$ and k estimates of $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}$, each using l samples of $\boldsymbol{\xi}_1$.

(ii) For each estimate, solve (KS) for s different uncertainty set sizes, yielding $k \times s$ solutions.

(iii) For each solution, generate n samples of $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ from (4.9) to probe constraint satisfaction.

(iv) Average over $k \times s$ objective values and $k \times s$ fractions of constraint satisfaction.

In module (i), let $l = 500$ and $k = 30$. Sample averages are used to estimate $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}$. The coefficients $\mathbf{c}_1$ and $\mathbf{c}_2$ are sampled from normal distributions centered at $\mathbf{e}$ and $1.25 \times \mathbf{e}$, respectively with covariance matrix $\frac{1}{100}\boldsymbol{\Sigma}$. For the constraint RHS, the budget is $B = 40$.
In (ii), the experiment is conducted for $s = 20$ different uncertainty set sizes $r_1 = r_2 \in [0, 4]$.
In (iii), the constraint satisfaction is measured using $n = 500$ samples.

To probe each setting, the knapsack problem is solved for both CU and NC sets. This is parallel to the difference between models in Figure 4.1(b) and (d) of the introductory Example 2. Figure 4.2 shows the average constraint satisfaction (left) and the average objective value

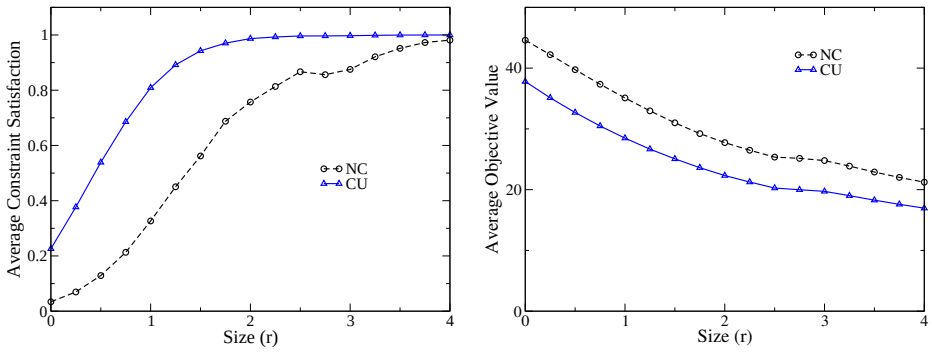


Figure 4.2. Comparison of connected and non-connected sets for the robust knapsack problem at different set sizes: (left) fraction of constraint satisfaction, and (right) objective value.

(right) for a varying size r of the uncertainty set. Figure 4.3 shows the average objective value for any level of constraint satisfaction (left) and show how the number of non-zero allocations

change as the set size increases (right). The data points at $r_1 = r_2 = 0$ correspond to the nominal problem with no uncertainty.

We now summarize our observations.

- ***Effect of Uncertainty Set Size***: For both models, constraint satisfaction increases and average objective value decreases with r. The average objective is lower for CU, since connectedness magnifies the worst-case. Note that the objective value is only measured, if constraints are satisfied.
- ***CU vs. NC***: For any r, CU solutions have higher constraint satisfaction than NC solutions. This is because connected sets account for dependency on the first period and provide additional protection beyond that of NC sets. Though this benefit comes with lower objective value, for any level of satisfaction, the average objective of CU is better than NC (see Figure 4.3 left).
- ***First vs Second period solutions***: For a single estimate in (ii), Figure 4.3 (right) shows that the optimal solution gradually concentrates only on $\mathbf{x}_1$ for CU and only on $\mathbf{x}_2$ for NC as r increases. For NC, this is because $\mathbf{c}_2$ tends to be higher. For CU, the second period weights are magnified due to connectedness, without corresponding benefits. Overall more components of $\mathbf{x}_1$ and $\mathbf{x}_2$ are non-zero for NC, resulting in higher objective but worse constraint satisfaction and vice versa.
- ***Negative Correlation***: When consecutive uncertainties are negatively correlated, CU sets achieve lower constraint satisfaction but higher objective value than NC sets at any r (see Appendix A.4). This is because the worst-case uncertainty is dampened for CU sets by the negative correlation. For any level of constraint satisfaction, both models perform similarly. The solutions concentrate on $\mathbf{x}_2$ for increasing r because of higher $\mathbf{c}_2$ (see the figures in A.4).

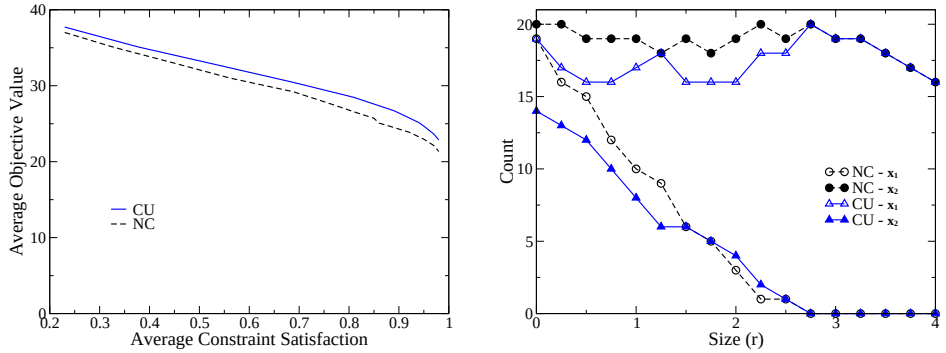


Figure 4.3. Comparison of connected and non-connected sets: (left) objective vs constraint satisfaction, and (right) for a single iteration, the number of non-zero variables vs set size.

Summary

CU sets exhibit clear advantages over non-connected sets for positively autocorrelated uncertainties. Though CU has lower objective function values, it outperforms NC for any given level of constraint satisfaction. The numerical analysis of modeling with CU sets responds to the research question *Q2*, highlighting improvements in constraint satisfaction of popular uncertainty sets. Note that an alternative NC model with growing uncertainty sets over time, as displayed in Figure 4.1c, also does not improve over CU. Therefore, we recommended to model with CU sets, particularly when uncertainties are positively correlated.

4.5. DRO Application: Portfolio Optimization

The area of portfolio optimization has seen sizable advancements by systematically managing uncertainty, and DRO lends itself as a natural modeling framework. In particular, different types of ambiguity sets have been proposed, such as moment constrained sets [35, 79] and sets bounded by the Wasserstein metric [39] etc. In this section, we evaluate the performance of our CU framework on portfolio optimization problems, resembling those by Delage and Ye [35]. We investigate two instances of this problem.

Multi-asset Portfolio Problem. Using historical data on 5 stocks, we probe our approach in a real-world setting. We model the asset returns with a time series fitted to the historic data for two periods of one week each. We solve this problem for randomly selected days to evaluate average performance on actually realized returns. We compare the CU model to two competing DRO model: one with same ambiguity sets for both periods, and one with the second period ambiguity set centered at the expected second period return (unconditioned). We observe that the standard deviation of wealth for the CU model is lower than both DRO models, while achieving comparable returns. This is because the CU model captures the compounding worst-case effects of autocorrelation, while managing conservatism. We relegate the numerical details to the Appendix A.5.

Two-asset Portfolio Problem. To gain further intuition in a controlled environment, we also simulate a two-asset portfolio optimization problem with synthetic data. The goal of this experiment is to probe the long term benefits of anticipating the behavior of uncertainties when making decisions as opposed to short term decision-making based on a fixed uncertainty model. Specifically, we consider a 2-period problem, in which the choice of the portfolio has to be made at the beginning. Such a problem can be expressed as

$$\begin{aligned}
&\max_{\mathbf{x}_1,\mathbf{x}_2} \inf_{P_1\in\widetilde{\mathcal{U}}_1} \mathbb{E}_{P_1}\left[u_1(\mathbf{x}_1,\boldsymbol{\xi}_1) + \inf_{P_2\in\widetilde{\mathcal{U}}_2(\xi_1)} \mathbb{E}_{P_2}[u_2(\mathbf{x}_2,\boldsymbol{\xi}_2)]\right] \\
&\text{s.t.} \quad \mathbf{e}^\top\mathbf{x}_t = 1 && \forall t = 1,2 \\
&\qquad \mathbf{x}_t \geq \mathbf{0} && \forall t = 1,2,
\end{aligned}$$

where u_t are utility functions, $\mathbf{x}_t$ decision variables, $\boldsymbol{\xi}_t$ return realizations, and $\widetilde{\mathcal{U}}_t$ the respective distributional uncertainty sets. The uncertainty sets $\widetilde{\mathcal{U}}_1$ and $\widetilde{\mathcal{U}}_2(\boldsymbol{\xi}_1)$ are as defined in (D). Each period corresponds to one week, and at the end of the first week, the assets can be reassigned. However, the reallocation decision has to be specified at the beginning when solving the problem.

To illustrate the impact of correlation among assets and correlation over time, consider the above problem with a concave utility function for each period given by $\min(1.5r, 0.015+r, 0.06+0.2r)$. Here, $r = x_1\xi_{11} + x_2\xi_{12}$ represents the portfolio return in each period. For this problem, the expected return in the first period resides within a box centered at $\boldsymbol{\mu}_1 = [0.03, 0.06]$ with size $\boldsymbol{\delta} = [0.02, 0.02]$. The expected return in the second period lies inside a similar set whose center depends on the realized return of the first period as $\boldsymbol{\mu}_1 + \omega \cdot (\boldsymbol{\xi}_1 - \boldsymbol{\mu}_1)$, where ω is a parameter controlling the correlation over time. Furthermore, the covariance of the return in both periods is bounded by $\boldsymbol{\Sigma}_1$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.005 & 0.005 \cdot \rho \\ 0.005 \cdot \rho & 0.005 \end{bmatrix},$$

where ρ measures the correlation among the asset returns. We vary the parameters ω over the range $[-2, 2]$ and $\rho \in [-1, 1]$ to study the impact of correlations on the following metrics of the portfolio model:

(1) *Asset allocation:* to observe the behavior of decisions and to develop intuition (Figures 4.4).

(2) *Average realized wealth:* to test the performance under realistic (random) setting (Figure 4.5).

(3) *Difference in standard deviation of the CU and DRO models:* to probe sensitivity (Figure 4.6).

We now discuss the outcomes of the numerical experiments with regard to these metrics. Additional metrics are discussed in Appendix A.6. In each of the upcoming figures, the horizontal axis represents the correlation over time ω, and the vertical axis the correlation among the assets. We will divide the discussion into these two aspects. The color in the plots represents the metric under consideration and its values are displayed in the legend.

4.5.1. Asset Allocation

Figure 4.4 displays the asset allocation for Asset 1 for the CU model for period 1 (left), period 2 (center), and the DRO model in period 1 (right), which is the same as period 2. The sum of allocations is one. Figure 4.4 displays that both the CU and DRO models allocate more wealth to Asset 2. This is due to the fact that Asset 2 has higher expected returns.

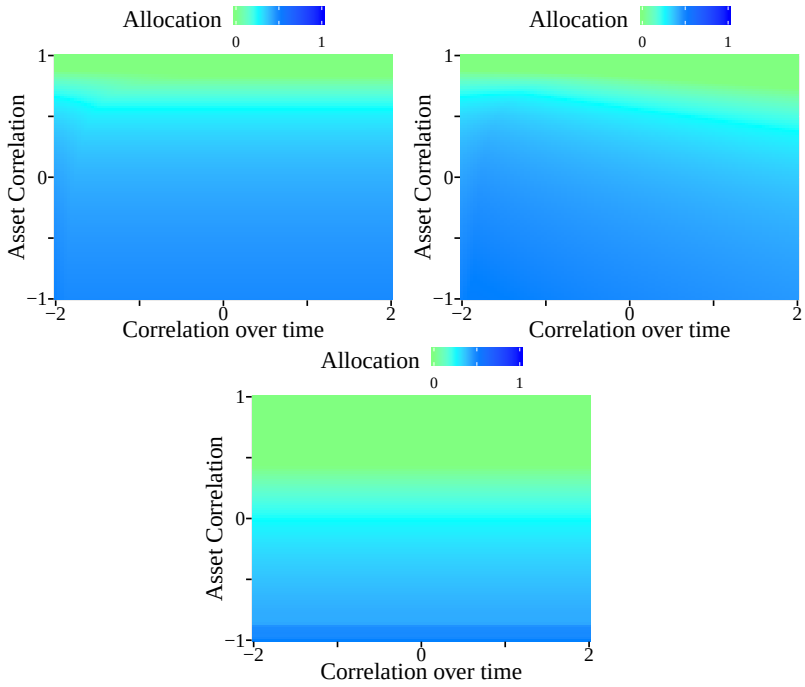


Figure 4.4. CU allocation to Asset 1 for CU period 1 (left), CU period 2 (center), and DRO period 1 (right).

Asset Correlation. For both CU and DRO models, the allocation to Asset 2 is reduced if the assets are positively correlated. This is because a positive correlation reduces the protection against the worst case that would arise by spreading the wealth among the two assets. A negative correlation among the assets increases this protection, resulting in a broader spread.

Correlation over time. We also observe that as the correlation over time increases, the CU model concentrates on Asset 2. This is because it has higher expected return and the benefit

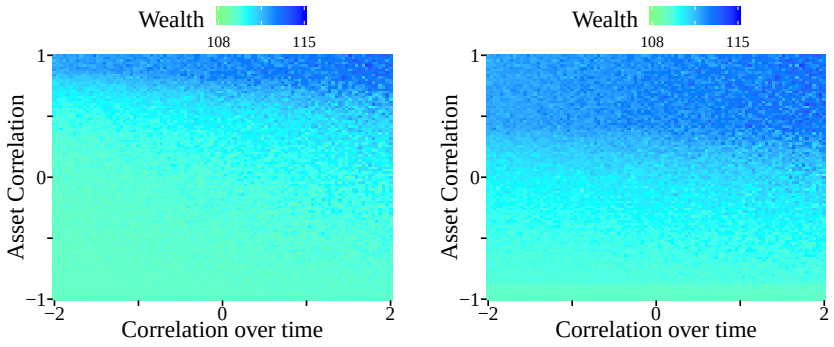


Figure 4.5. Average realized wealth for CU and DRO at the end of period 2.

of hedging against bad returns by spreading the wealth among both assets is reduced by the positive correlation. The correlation over time mostly affects the assets in the second period. A higher correlation leads to more allocations to Asset 2 since it has a higher expected value, and thus leads to better worst-case performance than Asset 1.

4.5.2. Average wealth

Asset Correlation. Figure 4.5 shows the average wealth at the end of the second period. It can be observed that high correlation between the assets leads to higher wealth. This contrasts with the worst-case performance and is because positive correlations also promote high realized returns. The wealth increases faster for the DRO model due to lower conservatism of the model.
Correlation over time. The realized wealth increases with correlation over time as it leads to higher realized returns. This behavior is more visible for the CU as compared to the DRO model.

4.5.3. Difference in CU and DRO models

Figure 4.6 (left) shows the difference in wealth standard deviation between the CU and DRO models and Figure 4.6 (right) the difference in worst-case wealth between these two models.

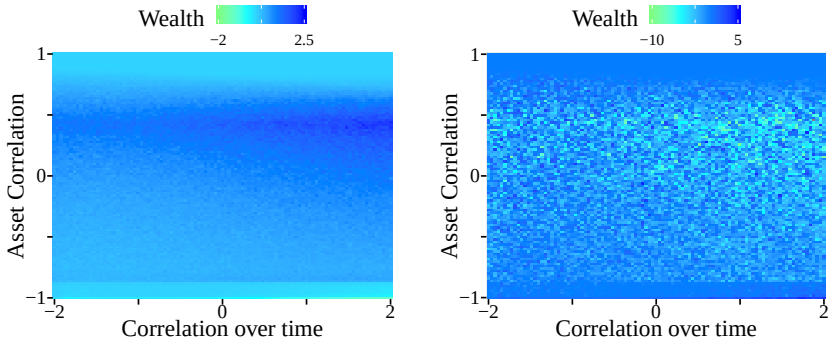


Figure 4.6. Difference in wealth standard deviation (left) and difference in worst-case wealth (right) for period 2.

Asset Correlation. The difference in standard deviations is always positive, increases with the asset correlation, and achieves a peak at 0.4. That means the CU model always leads to lower standard deviation, highlighting robustness of solutions. Also, the worst-case wealth of the CU model tend to perform better as the correlation increases and achieving best performance at 0.4.

Correlation over time. The benefits of the CU model in terms of lower standard deviation and better worst-case wealth increases with correlation over time.

4.6. Conclusion

In multistage optimization under uncertainty, the model of the underlying uncertainty plays a decisive role for both the solution quality as well as the computational effort. Current techniques usually assume a collection of predetermined sets which are independent of each other. Given that in many applications the uncertainty is connected over multiple periods, these methods are either over conservative or do not adequately capture all realizations of the uncertainty. We introduce a new modeling framework in which the uncertainty model depends on past realizations. This chapter extends the efficacy of robust optimization as well as distributionally robust optimization to the connected uncertainty paradigm. We study commonly used constraints and uncertainty set dependencies and provide their respective reformulations. Specifically, for robust settings, tractable reformulations are developed for polyhedral and ellipsoidal uncertainty sets

with linear and quadratic dependence, as they occur in a wide range of decision-making processes. When the uncertainty is modeled in a distributionally robust fashion, we provide reformulations for moment constrained ambiguity sets with the mean depending on the past, as in time series settings. Numerical experiments on a knapsack problem exhibit sizable improvements in constraint satisfaction and better objective performance for any fixed level of constraint satisfaction, when uncertainties are modeled with the proposed connected sets. Similarly, a distributionally robust portfolio optimization problem achieves approximately the same expected returns at narrower standard deviations, which highlights the robustness of solutions, when modeling with CU sets. Thus, we recommend using connected uncertainty sets for autocorrelated or time-varying uncertainties because they improve the performance over non-connected sets. Since in multi-period problems, uncertainties are naturally connected across periods, the proposed approach offers a general modeling framework that can be applied to numerous operational applications.

In the next chapter, we focus on an application of connected uncertainty sets to a classification problem. Specifically, we leverage CU sets to extend the Minimax Probability Machine model to develop policies for adaptive classifying surfaces.

CHAPTER 5

Robust Classifiers for Streaming Data

5.1. Introduction

When dividing data into two or more groups, classification methods have enhanced the benefits of machine learning in many application domains such as image recognition [75], spam filters [36], and medical diagnosis [70], amongst others. Two common approaches for classification are estimating a linear or non-linear surface that separates the classes (e.g., support vector machines [92], maximum probability machines [67], and decision trees [28]) and predicting the probability of a data point belonging to either class (e.g., neural networks [86], logistic regression [53]). These two directions have also been combined in integrated approaches such as ensemble learning methods [97].

In general, classification models assume that the distribution of each class is constant over time, enabling a time invariant classifier. In many applications, however, the underlying distributions can vary with time. For example, spammers change their tactics in response to detection algorithms, leading to a change in the observed data [36]. Similarly, in credit fraud detection, the characteristics of fraud change over time to counter the actions taken by banks or to adjust to the economy [98]. Using static models for such dynamically changing data leads to classifiers whose performance degrades over time.

To address these issues, modifications of existing algorithms are proposed which adapt to new data, such as time-windows [69, 80, 107], forgetting methods [5, 61], and incremental learning approaches [36, 102]. Most of these methods do not impose assumptions on how the data varies over time. Despite their broad applicability, these approaches have limited capability to leverage

a potentially known structure in the data in order to improve predictions. Recently, Kumagai and Iwata incorporated time-series and Gaussian processes into the classification problem while fitting statistical models [63, 64, 65].

In this chapter, we develop three classification methods that capture the changing behavior of each class. Specifically, we consider data which can be modeled by time series and leverage this structure when computing the classification surface. This allows the classification surface to adapt to the newly observed data at each step, transforming into a new surface that optimally reflects the changing behavior of data, and overcoming the limitations of static classifiers and improving the accuracy. We assume the time series predicts the mean and covariance of the next period based on observed data. This exposes the classifier to two layers of uncertainty: (i) the moment information from the time series models inherits uncertainty on the underlying distribution; and (ii) updates in time series models depend on past realizations, which are a-priori uncertain, when devising the classifier.

In order to account for potential distributional uncertainty, as in layer (i), we employ a Minimax Probability machine (MPM) model, as introduced by Lanckriet et al. [67]. The MPM model finds the surface that maximizes the probability of correct classification, while accounting for distributional uncertainty in the classes by the use of distributionally robust optimization and has been used in several applications [33, 40, 49]. Since the true class distributions are unknown, the MPM model optimizes for the worst-case distribution among all possible distributions for a given mean and covariance. In order to account for uncertainty in time series model, as in layer (ii), we use the method of robust optimization [12]. Its paradigm posits a *robust optimal solution* that is feasible for all realizations of the uncertainty and optimal for the worst. To ensure this, the original constraints are replaced by their *robust counterpart* which protects against infeasibility. The practical advantage of this approach is that it does not require detailed information on the uncertainty and rather describes its structure by sets. The geometry of these uncertainty

sets determines the computational tractability [17]. In this chapter, the uncertain realizations are modeled to reside in ellipsoids, which depend on the time series. The goal is to find how to modify the classifier for every possible realization, including the worst-case. The resulting classifier will then be *robust* to both layers of uncertainty for each observation, and hence relevant to a broad range of streaming applications.

The topic of this chapter is similar to non-stationary online optimization with a variation budget [24, 25, 59]. In both problems the underlying distribution, that generates the data, changes over time. However, the key difference is how this change occurs. In the case of a variation budget, the sum of all inter-period changes is bounded. In our work, however, we use a time-series model to specify how the distribution will change over time. The non-stationary online optimization problems in the literature primarily focus on minimizing a regret function, while our model directly optimizes the objective as it has no restrictions on the data generating distribution. These two aspects distinguish our work from the non-stationary online optimization problem. More specifically, our main contributions in this chapter can be summarized as following:

(1) We develop the *Adaptable Robust Classifier* (AdRC), where the time series predicts the next period mean and covariance, which are used to solve the MPM problem via a least squares algorithm. The resulting surface then classifies the incoming realization. Once the new data point is classified, it is included in the time series of its true class and the new moments are updated. This process is repeated for each new observation of the streaming data.

(2) In a more proactive approach, we develop the *Adjustable Robust Classifier* (AjRC), where the time series is directly imbedded into the MPM model in order to find a policy on how to update the classifying surface for each new data point. The advantage is that obtaining a new classifier does not require re-solving the full problem. However,

the policies require a semi-definite program, which can be computationally expensive for large data.

(3) To improve computation, we provide the *Approximate Adjustable Robust Classifier* (AAjRC), where the worst-case approach of AjRC is approximated by a second order conic program.

(4) We numerically evaluate the computational performance of these models on synthetic data and probe the impact of autocorrelation as well as the practical relevance of the methods for hourly wind speed data stream.

5.2. Model

In this section, we discuss the model components for the classifying algorithms with regard to the computation of the classifying surface via MPM and modeling the time varying nature of the data.

5.2.1. Minimax Probability Machine Model

The MPM problem assumes that data is generated from two classes X and Y with known means and covariances [67]. Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$ be random vectors in a binary classification problem with means $\boldsymbol{\mu}_x, \boldsymbol{\mu}_y$ and covariance matrices $\boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_y$. The MPM model determines a hyper plane characterized by $\mathbf{a}$ and b, separating these two classes with maximum probability over all distributions that have the given mean and covariance matrix. The model can be expressed as

$$
\begin{aligned}
&\max_{\alpha,\mathbf{a},b} \ \alpha \\
\text{(MPM)} \qquad &\text{s.t.} \inf_{\mathbb{P}_x \in \mathcal{M}_x} \mathbb{P}_x[\mathbf{a}^\top \mathbf{x} \geq b] \geq \alpha \\
&\qquad\quad \inf_{\mathbb{P}_y \in \mathcal{M}_y} \mathbb{P}_y[\mathbf{a}^\top \mathbf{y} \leq b] \geq \alpha,
\end{aligned}
$$

where $\mathcal{M}_x$ and $\mathcal{M}_y$ denote the set of distributions with means $\boldsymbol{\mu}_x, \boldsymbol{\mu}_y$ and covariances $\boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_y$. The probability of correct classification is denoted by α. This model can be tractably reformulated and solved using a least squares algorithm (see [67] for more details). The dual reformulation of (MPM) can be expressed as

$$\text{(D-MPM)} \qquad \begin{aligned} &\min_{r,\mathbf{u},\mathbf{v}} r \\ &\text{s.t. } \boldsymbol{\mu}_x + \Sigma_x^{\frac{1}{2}}\mathbf{u} = \boldsymbol{\mu}_y + \Sigma_y^{\frac{1}{2}}\mathbf{w} \\ &\|\mathbf{u}\|_2 \leq r \\ &\|\mathbf{w}\|_2 \leq r. \end{aligned}$$

This dual problem can be interpreted as follows. Consider 2 ellipsoids centered at $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ with covariance matrices $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$. The problem (D-MPM) seeks to find the smallest size r of the ellipsoids, at which they touch. Consequently, the common tangent passing through the point of contact of these ellipsoids constitutes the classifying surface between the two sets of (static) data.

The problem (MPM) can be used to find the optimal classifying surface, when the data are independent over time. In this setting, we can aggregate the observed data in order estimate the mean and covariance matrix. However, when data at a period depend on the realizations of the previous periods, the classifying surface from (MPM) renders inaccurate. This is because the mean or the covariance matrix of the next period can differ from that of the current period due to the time-dependence. Such dependence can be modeled via time series.

5.2.2. Time Series Model

The data is assumed to be generated in the following general way. At each classification time period t, either an element of class X is observed with probability p or an element of class Y

with probability $1-p$. If we observe an element of X, it is assumed to have been generated by the time series model

$$\mathbf{x} = \mathbf{a}_x + \mathbf{A}_{xx}\mathbf{x}_{\tau_x(t)} + \mathbf{A}_{xy}\mathbf{y}_{\tau_y(t)} + \boldsymbol{\epsilon}_{x,t}.$$

On the other hand, if an element of Y is observed, we assume it to be generated by

$$\mathbf{y} = \mathbf{a}_y + \mathbf{A}_{yx}\mathbf{x}_{\tau_x(t)} + \mathbf{A}_{yy}\mathbf{y}_{\tau_y(t)} + \boldsymbol{\epsilon}_{y,t}.$$

Here, $\tau_x(t)$ and $\tau_y(t)$ are look up functions for finding the time of the most recently (before t) observed elements of each class. The uncertain elements $\boldsymbol{\epsilon}_{x,t}$ and $\boldsymbol{\epsilon}_{y,t}$ follow unknown distributions with mean $\mathbf{0}$ and $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$. This dynamic is illustrated in Figure 5.1 for the fifth period of a data stream. The mean of the next period realization at t conditioned on the recent

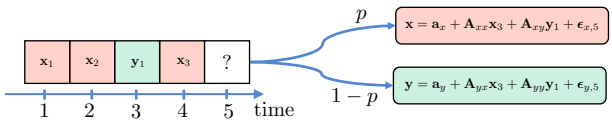


Figure 5.1. An illustration of a data stream of two classes X and Y for 5 time periods.

realizations $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ can be written as

$$\begin{aligned}
\boldsymbol{\mu}_x(\hat{\mathbf{x}}, \hat{\mathbf{y}}) &= \mathbf{E}[\mathbf{x}|\mathbf{x}_{\tau_x(t)} = \hat{\mathbf{x}}, \mathbf{y}_{\tau_y(t)} = \hat{\mathbf{y}}] \\
&= \mathbf{a}_x + \mathbf{A}_{xx}\hat{\mathbf{x}} + \mathbf{A}_{xy}\hat{\mathbf{y}} \\
\boldsymbol{\mu}_y(\hat{\mathbf{x}}, \hat{\mathbf{y}}) &= \mathbf{E}[\mathbf{y}|\mathbf{x}_{\tau_x(t)} = \hat{\mathbf{x}}, \mathbf{y}_{\tau_y(t)} = \hat{\mathbf{y}}] \\
&= \mathbf{a}_y + \mathbf{A}_{yx}\hat{\mathbf{x}} + \mathbf{A}_{yy}\hat{\mathbf{y}}.
\end{aligned} \tag{TS}$$

Note that this description of the time series model is general and captures the cross dependence of the two classes (via $\mathbf{A}_{xy}$ and $\mathbf{A}_{yx}$) over time. It also allows extensions to dependency structures on multiple previous realizations of either class.

5.3. Adaptable Robust Classifier: AdRC

AdRC uses the time series model to predict the next period mean and covariance matrix for each class. At each time period, the problem (MPM) is formulated as the following problem

$$\min_{\mathbf{u},\beta>0,\eta>0} \beta + \frac{1}{\beta}\|\boldsymbol{\Sigma}_x^{\frac{1}{2}}(\mathbf{a}_0 + \mathbf{F}\mathbf{u})\|_2^2 + \eta + \frac{1}{\eta}\|\boldsymbol{\Sigma}_y^{\frac{1}{2}}(\mathbf{a}_0 + \mathbf{F}\mathbf{u})\|_2,$$

which can be solved via an iterative least squares algorithm. Here, $\mathbf{a}_0 = (\boldsymbol{\mu}_x - \boldsymbol{\mu}_y)/\|\boldsymbol{\mu}_x - \boldsymbol{\mu}_y\|_2^2$ and $\mathbf{F}$ is a basis for the null space of $\boldsymbol{\mu}_x - \boldsymbol{\mu}_y$, i.e., $\mathbf{F}$ is an orthogonal matrix whose columns span the subspace of vectors orthogonal to $\boldsymbol{\mu}_x - \boldsymbol{\mu}_y$. Solving the reformulated problem iteratively yields the classifier $\mathbf{a} = \mathbf{a}_0 + \mathbf{F}\mathbf{u}$ with $b = \mathbf{a} \cdot \boldsymbol{\mu}_x - \beta/(\beta - \eta)$. This surface then classifies the incoming point of the stream, which is then used to update the estimate of the mean for both classes for the next period's classification. The steps of AdRC are summarized below.

Algorithm 1 The steps of AdRC for classifying streaming data.

1: Initialize: $t = 1$ and previous data points $\mathbf{x}_0$ and $\mathbf{y}_0$
2: **while** New streaming data point $\mathbf{z}_t \in \mathbb{R}^n$ is available **do**
3: Predict $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ using model (TS) and $\mathbf{x}_{\tau_x(t)}, \mathbf{y}_{\tau_y(t)}$.
4: Solve problem (MPM) using iterative least squares.
5: Use generated classifier to classify $\mathbf{z}_t$.
6: Record the observed true class for $\mathbf{z}_t$.
7: $t = t + 1$
8: **end while**

The AdRC requires repeated computation and becomes expensive for large-scale data streams. Since the AdRC method consists of two separable steps of (i) predicting the next period moments and (ii) solving (MPM) using these moments, the improvement of each step can be directly beneficial. For example, we can use better prediction methods for the moments or map the data

into higher dimensions, also known as kernelization [67]. Nevertheless, the aggregating aspect of the repeated computation for AdRC imposes limitations.

5.4. Adjustable Robust Classifier: AjRC

To overcome the repeated computation, we propose a method to solve the adjustable reformulation of problem (D-MPM). The resulting policy prescribes how to adjust the classifier for newly observed data from the stream. For this, we reformulate the problem (D-MPM) to allow the variables to be functions of the uncertain realizations of previous periods. These variables predict the contact points of the two ellipsoids as their centers vary with past realizations. This policy is obtained by minimizing the worst-case norm of the contact point for all possible realizations from the previous period. Given the stationarity of time series, we model previous realizations to reside in ellipsoidal sets centered around the unconditional means. Therefore, the adjustable dual MPM can be given by

$$
\begin{aligned}
&\min_{\substack{r,\mathbf{u}(\mathbf{x},\mathbf{y})\\ \mathbf{w}(\mathbf{x},\mathbf{y})}} \max_{\substack{\mathbf{x}\in\mathcal{U}_x\\ \mathbf{y}\in\mathcal{U}_y}} r(\mathbf{x},\mathbf{y}) \\
&\qquad \text{s.t. } \boldsymbol{\mu}_x(\mathbf{x},\mathbf{y}) + \boldsymbol{\Sigma}_x^{1/2}\mathbf{u}(\mathbf{x},\mathbf{y}) = \boldsymbol{\mu}_y(\mathbf{x},\mathbf{y}) + \boldsymbol{\Sigma}_y^{1/2}\mathbf{w}(\mathbf{x},\mathbf{y}) \\
&\qquad\qquad \forall \mathbf{x}\in\mathcal{U}_x \ \forall \mathbf{y}\in\mathcal{U}_y \\
&\qquad \|\mathbf{u}(\mathbf{x},\mathbf{y})\|_2 \le r(\mathbf{x},\mathbf{y}) \quad \forall \mathbf{x}\in\mathcal{U}_x,\ \mathbf{y}\in\mathcal{U}_y \\
&\qquad \|\mathbf{w}(\mathbf{x},\mathbf{y})\|_2 \le r(\mathbf{x},\mathbf{y}) \quad \forall \mathbf{x}\in\mathcal{U}_x,\ \mathbf{y}\in\mathcal{U}_y.
\end{aligned}
\tag{AD-MPM}
$$

The uncertainty sets $\mathcal{U}_x$ and $\mathcal{U}_y$ represent the unconditioned (independent of past) sets of $\mathbf{x}$ and $\mathbf{y}$. For tractability, we make the following common assumption.

Assumption 5.4.1. *The uncertainty sets $\mathcal{U}_x$ and $\mathcal{U}_y$ are ellipsoids with centers and covariance matrices independent of the realizations of the time series of classes X and Y.*

Given Assumption 5.4.1, we model the corresponding uncertainty sets as

$$\begin{aligned}\mathcal{U}_x &= \left\{\mathbf{x}|\mathbf{x} = \boldsymbol{\mu}_x^0 + \mathbf{L}_x\mathbf{u} : \|\mathbf{u}\|_2 \leq \kappa_x\right\} \\ \mathcal{U}_y &= \left\{\mathbf{y}|\mathbf{y} = \boldsymbol{\mu}_y^0 + \mathbf{L}_y\mathbf{w} : \|\mathbf{w}\|_2 \leq \kappa_y\right\},\end{aligned}$$

where $\mathbf{L}_x\mathbf{L}_x^\top = \boldsymbol{\Sigma}_x$ and $\mathbf{L}_y\mathbf{L}_y^\top = \boldsymbol{\Sigma}_y$ and $\boldsymbol{\mu}_x^0$ and $\boldsymbol{\mu}_y^0$ are the unconditional means for (TS). Problem (AD-MPM) is an adjustable robust optimization (ARO) problem, which is known to be NP-hard [11]. This can be overcome by approximating the functions $\mathbf{u}(\mathbf{x}, \mathbf{y}), \mathbf{w}(\mathbf{x}, \mathbf{y})$ and $r(\mathbf{x}, \mathbf{y})$ with affine or static decision rules.

5.4.1. Affine Approximation

Affine decision rules, in addition to improving tractability, also maintain good approximation properties for both the average and worst-case scenarios. Consider the time series models $\boldsymbol{\mu}_x(\mathbf{x}, \mathbf{y})$ and $\boldsymbol{\mu}_y(\mathbf{x}, \mathbf{y})$ in (TS) and define $\boldsymbol{\xi} = (1, \mathbf{x}^\top, \mathbf{y}^\top)^\top$. Then we can write

$$\begin{aligned}\boldsymbol{\mu}_x(\boldsymbol{\xi}) &= [\mathbf{a}_x \mid \mathbf{A}_{xx} \mid \mathbf{A}_{xy}]\,\boldsymbol{\xi} = \overline{\mathbf{X}}\boldsymbol{\xi} \\ \boldsymbol{\mu}_y(\boldsymbol{\xi}) &= [\mathbf{a}_y \mid \mathbf{A}_{yx} \mid \mathbf{A}_{yy}]\,\boldsymbol{\xi} = \overline{\mathbf{Y}}\boldsymbol{\xi}.\end{aligned}$$

Similarly, we express the functional variables $\mathbf{u}(\mathbf{x}, \mathbf{y})$ and $\mathbf{w}(\mathbf{x}, \mathbf{y})$ as the following affine functions

$$\begin{aligned}\mathbf{u}(\mathbf{x}, \mathbf{y}) &= \mathbf{u}_0 + \mathbf{U}_x\mathbf{x} + \mathbf{U}_y\mathbf{y} \\ &= \mathbf{u}(\boldsymbol{\xi}) = [\mathbf{u}_0 \mid \mathbf{U}_x \mid \mathbf{U}_y]\boldsymbol{\xi} = \overline{\mathbf{U}}\boldsymbol{\xi}\end{aligned}$$

$$\begin{aligned}\mathbf{w}(\mathbf{x}, \mathbf{y}) &= \mathbf{w}_0 + \mathbf{W}_x\mathbf{x} + \mathbf{W}_y\mathbf{y} \\ &= \mathbf{w}(\boldsymbol{\xi}) = [\mathbf{w}_0 \mid \mathbf{W}_x \mid \mathbf{W}_y]\boldsymbol{\xi} = \overline{\mathbf{W}}\boldsymbol{\xi}\end{aligned}$$

$$\mathbf{r}(\mathbf{x}, \mathbf{y}) = r,$$

where $\mathbf{u}_0, \mathbf{w}_0, \mathbf{U}_x, \mathbf{W}_x, \mathbf{U}_y$ and $\mathbf{W}_y$ are the optimization variables and r is the static approximation of the radius. We also redefine the uncertainty sets $\mathcal{U}_x$ and $\mathcal{U}_y$ as a joint uncertainty set over $\boldsymbol{\xi}$ through $\mathcal{U} = 1 \times \mathcal{U}_x \times \mathcal{U}_y$. Then, the problem (AD-MPM) with affine decision rules can be expressed as

$$\text{(AjRC)} \qquad \begin{aligned} \min_{r, \overline{\mathbf{U}}, \overline{\mathbf{V}}} \; & r \\ \text{s.t. } & \overline{\mathbf{X}}\boldsymbol{\xi} + \boldsymbol{\Sigma}_x^{\frac{1}{2}}\overline{\mathbf{U}}\boldsymbol{\xi} = \overline{\mathbf{Y}}\boldsymbol{\xi} + \boldsymbol{\Sigma}_y^{\frac{1}{2}}\overline{\mathbf{W}}\boldsymbol{\xi} && \forall \boldsymbol{\xi} \in \mathcal{U} \\ & \|\overline{\mathbf{U}}\boldsymbol{\xi}\|_2 \leq r && \forall \boldsymbol{\xi} \in \mathcal{U} \\ & \|\overline{\mathbf{W}}\boldsymbol{\xi}\|_2 \leq r && \forall \boldsymbol{\xi} \in \mathcal{U}. \end{aligned}$$

Note that the problem (AjRC) is a second order conic program with infinitely many constraints, making it difficult to solve. Let $\mathbf{f}(\mathbf{u}, \mathbf{U}_x, \mathbf{U}_y) = \mathbf{u} + \mathbf{U}_x\boldsymbol{\mu}_x^0 + \mathbf{U}_y\boldsymbol{\mu}_y^0$. Using the S-lemma, the following theorem reformulates this problem as a semi-definite program.

Theorem 5.4.1. *The robust counterpart of the problem* (AjRC) *is given by*

$$
\begin{aligned}
&\min_{r,\boldsymbol{\lambda},\overline{\mathbf{U}},\overline{\mathbf{W}}} \; r \\
&\quad \text{s.t. } \mathbf{a}_x + \boldsymbol{\Sigma}_x^{\frac{1}{2}}\mathbf{u}_0 = \mathbf{a}_y + \boldsymbol{\Sigma}_y^{\frac{1}{2}}\mathbf{w}_0 \\
&\qquad \mathbf{A}_{xx} + \boldsymbol{\Sigma}_x^{\frac{1}{2}}\mathbf{U}_x = \mathbf{A}_{yx} + \boldsymbol{\Sigma}_y^{\frac{1}{2}}\mathbf{W}_x \\
&\qquad \mathbf{A}_{xy} + \boldsymbol{\Sigma}_x^{\frac{1}{2}}\mathbf{U}_y = \mathbf{A}_{yy} + \boldsymbol{\Sigma}_y^{\frac{1}{2}}\mathbf{W}_y \\
&\begin{bmatrix}
r - \lambda_{xx}\kappa_x^2 - \lambda_{xy}\kappa_y^2 & \mathbf{f}(\mathbf{u}_0, \mathbf{U}_x, \mathbf{U}_y)^\top & \mathbf{0} & \mathbf{0} \\
\mathbf{f}(\mathbf{u}_0, \mathbf{U}_x, \mathbf{U}_y) & r\mathbf{I} & \mathbf{U}_x\mathbf{L}_x & \mathbf{U}_y\mathbf{L}_y \\
\mathbf{0} & \mathbf{L}_x^\top\mathbf{U}_x^\top & \lambda_{xx}\mathbf{I} & \mathbf{0} \\
\mathbf{0} & \mathbf{L}_y^\top\mathbf{U}_y^\top & \mathbf{0} & \lambda_{xy}\mathbf{I}
\end{bmatrix} \succeq 0 \\
&\begin{bmatrix}
r - \lambda_{yx}\kappa_x^2 - \lambda_{yy}\kappa_y^2 & \mathbf{f}(\mathbf{w}_0, \mathbf{W}_x, \mathbf{W}_y)^\top & \mathbf{0} & \mathbf{0} \\
\mathbf{f}(\mathbf{w}_0, \mathbf{W}_x, \mathbf{W}_y) & r\mathbf{I} & \mathbf{W}_x\mathbf{L}_x & \mathbf{W}_y\mathbf{L}_y \\
\mathbf{0} & \mathbf{L}_x^\top\mathbf{W}_x^\top & \lambda_{yx}\mathbf{I} & \mathbf{0} \\
\mathbf{0} & \mathbf{L}_y^\top\mathbf{W}_y^\top & \mathbf{0} & \lambda_{yy}\mathbf{I}
\end{bmatrix} \succeq 0 \\
&\qquad \lambda_{xx}, \lambda_{xy}, \lambda_{yx}, \lambda_{yy} \geq 0.
\end{aligned}
$$

The proof is in appendix A.7. This robust counterpart can be solved in polynomial time through the use of commercially available solvers. As a result, data streaming classifiers can be computed efficiently.

5.4.2. Estimating the classifying surface

The problem (AjRC) is similar to (D-MPM). It predicts the contact point of the two ellipsoids as their centers vary while minimizing the maximum norm (over all uncertain $\boldsymbol{\xi}$) of the normalized contact point. However, there is a key distinction from the problem (D-MPM), namely the

mean is known before computing the classifier and the solution occurs when the two ellipsoids touch. At this point, the classifying surface is the common tangent. However, this is not the case for (AjRC), because it calculates an approximate policy with a static radius. Consequently, the contact point from (AjRC) can be a point of intersection of the two ellipsoids. Since the static r is only for the worst-case, the actual radius for $\boldsymbol{\xi}$ can differ. Given realizations $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, let the two ellipsoids be expressed by

$$\mathcal{E}_1(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = \left\{ \mathbf{x} \mid (\mathbf{x} - \boldsymbol{\mu}_x(\hat{\mathbf{x}}, \hat{\mathbf{y}}))^\top \boldsymbol{\Sigma}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}_x(\hat{\mathbf{x}}, \hat{\mathbf{y}})) \leq r^2 \right\}$$
$$\mathcal{E}_2(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = \left\{ \mathbf{y} \mid (\mathbf{y} - \boldsymbol{\mu}_y(\hat{\mathbf{x}}, \hat{\mathbf{y}}))^\top \boldsymbol{\Sigma}_y^{-1} (\mathbf{y} - \boldsymbol{\mu}_y(\hat{\mathbf{x}}, \hat{\mathbf{y}})) \leq r^2 \right\}.$$

The solution of (AjRC) provides a point $\mathbf{v} = \mathbf{x} = \mathbf{y}$. The tangents of $\mathcal{E}_1(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ and $\mathcal{E}_2(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ at $\mathbf{v}$ are

$$\begin{aligned} \mathbf{g}_1 &= 2\boldsymbol{\Sigma}_x^{-1}(\mathbf{v} - \boldsymbol{\mu}_x(\hat{\mathbf{x}}, \hat{\mathbf{y}})) = 2\boldsymbol{\Sigma}_x^{-1}\mathbf{L}_x\mathbf{u} \\ \mathbf{g}_2 &= 2\boldsymbol{\Sigma}_y^{-1}(\mathbf{v} - \boldsymbol{\mu}_y(\hat{\mathbf{x}}, \hat{\mathbf{y}})) = 2\boldsymbol{\Sigma}_y^{-1}\mathbf{L}_y\mathbf{w}. \end{aligned}$$

The tangential surfaces at these points are given by $\mathbf{g}_1^\top \mathbf{x} = c_1$ and $\mathbf{g}_2^\top \mathbf{y} = c_2$ for some c_1 and c_2. Since both surfaces pass through the point $\mathbf{v}$, we can estimate c_1 and c_2 by $c_1 = \mathbf{g}_1^\top \mathbf{v}$ and $c_2 = \mathbf{g}_2^\top \mathbf{v}$. Note that the surfaces $\mathbf{g}_1, c_1$ and $\mathbf{g}_2, c_2$ are affine functions of realizations $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$.

Choosing the correct surface: We evaluate the probability of correct classification for both surfaces and choose the one with higher probability. Given the prior probability of class X and Y as p and $1 - p$ (see Figure 5.1), we can express the probability of correct classification as

$$\mathbb{P}(\mathbf{g}_i) = p \cdot \mathbb{P}(\mathbf{g}_i^\top \mathbf{x} \geq c_i) + (1 - p) \cdot \mathbb{P}(\mathbf{g}_i^\top \mathbf{y} \leq c_i).$$

Here, $\mathbb{P}(\mathbf{g}_i^\top \mathbf{x} \geq c_i)$ can be calculated assuming $\mathbf{x}$ to be normally distributed as $\mathcal{N}(\boldsymbol{\mu}_x(\hat{\mathbf{x}}, \hat{\mathbf{y}}), \boldsymbol{\Sigma}_x)$, or for the worst-case as $\mathbb{P}(\mathbf{g}_i^\top \mathbf{x} \geq c_i) = d^2/(1+d^2)$ with $d^2 = \max(\mathbf{g}_i^\top \boldsymbol{\mu}_x - c_i, 0)^2 / \sqrt{\mathbf{g}_i^\top \boldsymbol{\Sigma}_x \mathbf{g}_i}$.

The overall algorithm to obtain the AjRC classifier can be summarized as follows.

Algorithm 2 The steps of AjRC for classifying streaming data.

1: Initialize: $t = 1$ and previous data points $\mathbf{x}_0$ and $\mathbf{y}_0$
2: Solve the problem in Theorem 5.4.1 and obtain policy matrices $\overline{\mathbf{U}}$ and $\overline{\mathbf{W}}$.
3: **while** New streaming data point $\mathbf{z}_t$ is available **do**
4: Let $\boldsymbol{\xi}_t = (1, \mathbf{x}_{\tau_x(t)}, , \mathbf{y}_{\tau_y(t)})$.
5: Estimate contact point $\mathbf{v} = \overline{\mathbf{X}}\boldsymbol{\xi} + \mathbf{L}_x \overline{\mathbf{U}}\boldsymbol{\xi}$.
6: Calculate classifiers $\mathbf{g}_1 = 2\boldsymbol{\Sigma}_x^{-1}\mathbf{L}_x\overline{\mathbf{U}}\boldsymbol{\xi}$ and $\mathbf{g}_2 = 2\boldsymbol{\Sigma}_y^{-1}\mathbf{L}_y\overline{\mathbf{W}}\boldsymbol{\xi}$ with $c_i = \mathbf{g}_i^\top \mathbf{v}$.
7: Evaluate probabilities $\mathbb{P}(\mathbf{g}_i) = p \cdot \mathbb{P}(\mathbf{g}_i^\top \mathbf{x} \geq c_i) + (1-p) \cdot \mathbb{P}(\mathbf{g}_i^\top \mathbf{y} \leq c_i)$.
8: Use surface with higher probability to classify $\mathbf{z}_t$.
9: Record the observed true class of $\mathbf{z}_t$.
10: $t = t + 1$
11: **end while**

5.4.3. Approximate Adjustable Robust Classifier

The key challenge in efficiently solving the problem (AjRC) is its reformulation of the infinite-dimensional second order constraints. This difficulty can be alleviated by replacing the infinitely many constraints for the uncertainty set $\mathcal{U}$ with a finite number. In this context, we can express

the problem (AjRC) as

$$
\begin{aligned}
&\min_{r,\overline{\mathbf{U}},\overline{\mathbf{W}}} \; r \\
&\quad \text{s.t. } \mathbf{a}_x + \mathbf{L}_x\mathbf{u}_0 = \mathbf{a}_y + \mathbf{L}_y\mathbf{w}_0 \\
&\qquad \mathbf{A}_{xx} + \mathbf{L}_x\mathbf{U}_x = \mathbf{A}_{yx} + \mathbf{L}_y\mathbf{W}_x \\
&\qquad \mathbf{A}_{xy} + \mathbf{L}_x\mathbf{U}_y = \mathbf{A}_{yy} + \mathbf{L}_y\mathbf{W}_y \\
&\qquad \|\mathbf{u}_0 + \mathbf{U}_x\boldsymbol{\mu}_x^0 + \mathbf{U}_y\boldsymbol{\mu}_y^0 + \mathbf{U}_x\mathbf{L}_x\mathbf{u} + \mathbf{U}_y\mathbf{L}_y\mathbf{w}\|_2 \leq r \\
&\qquad\qquad \forall \mathbf{u} \in \hat{\mathbb{U}}, \mathbf{w} \in \hat{\mathbb{W}} \\
&\qquad \|\mathbf{w}_0 + \mathbf{W}_x\boldsymbol{\mu}_x^0 + \mathbf{W}_y\boldsymbol{\mu}_y^0 + \mathbf{W}_x\mathbf{L}_x\mathbf{u} + \mathbf{W}_y\mathbf{L}_y\mathbf{w}\|_2 \leq r \\
&\qquad\qquad \forall \mathbf{u} \in \hat{\mathbb{U}}, \mathbf{w} \in \hat{\mathbb{W}},
\end{aligned}
$$

where $\hat{\mathbb{U}}$ and $\hat{\mathbb{W}}$ denote a finite set of points to approximate the sets $\{\mathbf{u}|\|\mathbf{u}\|_2 \leq \kappa_x\}$ and $\{\mathbf{w}|\|\mathbf{w}\|_2 \leq \kappa_y\}$. This new problem is an SOCP, which can be solved significantly faster than the previous semi-definite program. This is crucial for large-scale data streams. Depending on the set of elements that are selected in $\hat{\mathbb{U}}$ and $\hat{\mathbb{W}}$, different approximations can be obtained. For example, we may obtain an inner approximation of the uncertainty set by choosing the vertices of an ellipsoid. Alternatively, we obtain an outer approximation by choosing the corners of a box enclosing the underlying ellipsoid.

5.5. Numerical Experiments

We conducted three experiments. An illustrative experiment serves to generate intuition for the classifying methods. The second experiment involves two sets of data with their means separated by increasing distances, in order to measure accuracy and speed. The last experiment

is a real-world test and uses meteorological data to predict high winds in the next hour. All codes are available online.

Computational Setup. The AdRC model is solved using the algorithm provided in [67], while the AjRC and AAjRC models are formulated as a SDP and a SOCP respectively. They are implemented using the JuMP library in the Julia programming language v1.0. The solvers used are Gurobi for the SOCP and Mosek for the SDP. They are solved on a machine with an Intel Core i7 processor with 32 GB RAM.

5.5.1. Illustrative Experiment

Consider the following two-dimensional example. Suppose the two time-series are given by

$$\mathbf{x}_t = [-0.4, -0.4] + 0.2\ \mathbf{x}_{\tau_x(t)} + 0.05\ \mathbf{y}_{\tau_y(t)} + \boldsymbol{\epsilon}_{x,t}$$

$$\mathbf{y}_t = [0, 0.15] + 0.05\ \mathbf{x}_{\tau_x(t)} + 0.2\ \mathbf{y}_{\tau_y(t)} + \boldsymbol{\epsilon}_{y,t}.$$

We initialize the series and discard the first 1000 points to achieve steady state behavior. The data generated after this warmup are used for classification, as illustrated in Figure 5.2 for four time periods. The circular data points represent class X, and the triangular ones, class Y. The

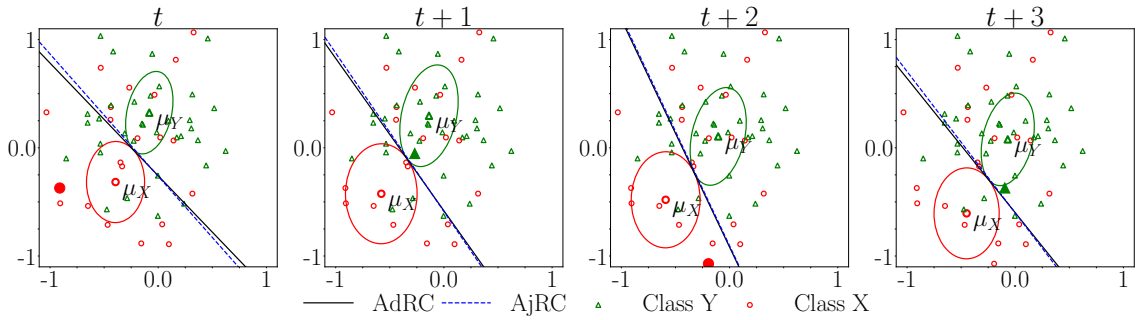


Figure 5.2. Four sequential iterations of classification. This segment is $[x_1, y_1, x_2, y_2]$.

incoming data point has a filled symbol. From Figure 5.2 we observe:

(1) for AdRC, the ellipsoids (centers have enlarged symbols) sizably change in each time period, causing changes to the classifier, and

(2) the AdRC (solid line) and AjRC (dashed) classifiers are close to each other, indicating that the affinely updated AjRC classifier closely predicts the AdRC classifier.

This proximity highlights the robustness of the AjRC surface to uncertainty in realized data.

5.5.2. Synthetic Data Stream Experiment

In a controlled environment, we measure the accuracy and speed of AdRC, AjRC, and AAjRC on synthetic data, while varying data size and the distance between the two class means. Specifically, for each dimension size, a time series generates the data. After an initial burn period of 1000 points, 4000 new data points are generated for the stream, which are divided into training and testing sets. The training sets are used to find an estimate of the time series in order to calculate the policy for updating the classifying surface and for predicting the next period's moments.

Table 5.1. Comparison of the three data stream classifiers on randomly generated time series.

Size	Distance	AdRC		AjRC		AAjRC	
		Accuracy	Time [s]	Accuracy	Time [s]	Accuracy	Time [s]
5	0.9	55.8	1.1	55.6	0.5	56.4	0.3
5	1.3	65.6	1.1	65.7	0.5	65.6	0.3
5	1.8	70.5	1.1	72.1	0.5	70.8	0.3
30	1.9	62.2	2.9	63.3	1558.4	62.9	4.9
30	2.8	67.4	2.6	69.3	1543.6	67.2	4.4
30	3.8	73.7	2.7	73.6	1753.8	73.5	4.0

Table 5.1 shows that the performance of all three models improves with the distance between the classes, highlighting the benefits of leveraging time series for well separated streams. We also

observe that the accuracy of the approximate model is comparable to the other two, indicating that the SOCP reformulation captures the effects well. While AjRC offers the benefit of a policy for classification, its computation for larger dimensions significantly degrades. However, AAjRC offers the same benefits as AjRC, while maintaining a comparable speed to AdRC. Depending on the application, this result allows practitioners to choose among the proposed methods.

5.5.3. Wind Speed Experiment

Predicting wind speeds plays an important role in aviation control. The hourly data stream from the weather station at the Chicago O'Hare Airport contains several predictors, such as temperature, air pressure, relative humidity, and past wind speeds. We divide the streaming dataset, from January 1, 2010 to June 30, 2010 from the National Climate Date Center [2], into two classes based on whether the wind speed was higher than the 90th percentile. We divide this data into a training set for the time series and a test set to measure the accuracy of the classifier. We compute the AdRC, AjRC, and AAjRC for this streaming data, assuming different levels of auto correlation in the time series.

Table 5.2. Accuracy of meteorology data stream classification for wind speeds.

Auto-Corr	AdRC	Time [s]	AjRC	Time [s]	AAjRC	Time [s]
3	63.4	2.4	63.0	2.5	58.3	0.8
4	64.2	2.3	64.4	6.6	62.0	0.8
5	66.1	2.3	66.4	15.3	58.7	0.8
6	67.4	3.7	67.0	37.4	62.2	3.4
7	68.5	2.2	68.4	61.5	64.3	0.9

Table 5.2 shows that higher auto correlations improve accuracy. Similar to the previous experiment, we observe that the computation of AjRC is more expensive, which can be overcome by AAjRC, while maintaining the accuracy. We observe that the probability approximation for

AjRC and AAjRC with a normal distribution leads to comparable results as assuming the worst case (not shown).

5.6. Conclusion

We extend the Minimax Probability Machine to classification of data streams. For this, the data variation over time is modeled by a cross class time series model. We present three extensions which provide classifiers for future observations. The adaptive method uses the time series model to predict the next period moments and solves the MPM problem to obtain the classifying surface. The adjustable method directly embeds the time series into the optimization problem and provides decision rules for adjusting the classifier to new observations. The approximate adjustable method reduces the problem size in order to manage larger data streams while maintaining the accuracy. We evaluate the performance of these models on numerical experiments and illustrate their benefits on synthetic and real-world data. These natural extensions are robust to distributional uncertainties and are suitable for general classification of data streams.

CHAPTER 6

Conclusion and Future Research

Robust Optimization has become a popular method for addressing optimization problems under uncertainty. So far, it has focused on uncertainty models that are fixed a priori. In this dissertation, we addressed the topic of variable uncertainty sets for robust optimization problems. These sets are affected by either the decisions made in the optimization problem, or by uncertainty realizations from previous time periods.

6.1. Decision Dependent Uncertainties

We introduced decision-dependent uncertainty sets and proved that robust optimization problems under such sets are NP-complete. We then introduced a special class of uncertainty sets via intersections of a box and a polyhedron and proved that their structure can be leveraged to improve the computation of these problems. This is achieved by moving the decision variables, affecting the uncertainty set, from the constraints defining the set, to the objective. This reformulates the robust counterpart as a convex optimization problem. These results were then illustrated through numerical experiments on a shortest path problem, highlighting the benefits that could be extracted by leveraging decision dependent uncertainty. They showed that the special class of uncertainty sets provides significant computational benefits over standard bilinear reformulations. The experiments also indicated that the benefits of uncertainty reduction significantly depend on the price for reducing the uncertainty. A low price can lead to large benefits, which decrease as the price increases. The exact impact of the reduction depends on the structure of the uncertainty set and the optimization problem.

We then extended this concept to allow for uncertainty in the influence of the decisions on the set. This influence uncertainty was modeled in both robust and stochastic optimization settings. We provided reformulations for both settings and presented results to simplify the robust setting for interval models of influence uncertainty. The impact of this uncertain influence was illustrated through numerical experiments on an electricity unit commitment problem with unknown load. We developed cut generation algorithms for this unit commitment problem with decision dependent uncertainty by leveraging the results from Chapter 2. We provided pre-computations of the worst-case scenarios for several of the unit commitment constraints. The numerical experiments displayed the impact of modeling the uncertain influence. Depending on the model, influence uncertainty led to differences in the decision to reduce the uncertainty. This indicates the importance of incorporating any uncertainty in the influence into the optimization problem.

6.2. Connected Uncertainties

In the second half, we studied the topic of connected uncertainty, where uncertainty realizations over multiple periods can affect each other. Specifically, we focused on a multi-period problem in which the uncertainty in one period can affect the uncertainty set of the future through a pre-specified model, such as a time series. We considered polyhedral and ellipsoidal uncertainty sets and computed the resulting reformulations, when the parameters of these sets are connected over time. We also extended these results to a distributionally robust optimization problem with moment uncertainty sets. Here, we assumed the first moment, i.e., the mean, to be dependent over time. This led to infinite dimensional optimization problems, which we conservatively approximated using static solutions. We evaluated the impact of such connections on a knapsack and a portfolio optimization problem. For the knapsack problem, we showed that the presence of connections can lead to more extreme worst-case scenarios, which are not

captured by standard models. CU models capture these scenarios and provide better feasibility at almost similar objectives. The experiments on the portfolio optimization problem displayed similar results with the CU model lowering the risk (standard deviation), while achieving similar wealth as the standard non-connected model.

We then leveraged the notion of connected uncertainty in the context of a minimax probability machine model to develop an adaptive classifier for streaming data. We assumed that the data of each class over time was modeled by a time series. This allowed us to develop two classifiers referred to as AdRC and AjRC. AdRC re-computed the classifying surface at each time step using the new predictions (from the time series) of the mean. AjRC was developed using a connected uncertainty model and allowed us to compute a policy for updating the classifier over time by solving a semi-definite optimization problem. We approximated AjRC as a second order conic program AAjRC to improve the computation of the optimal policy. These models were illustrated through experiments on synthetic and wind speed data. The results indicated that both models perform well, when there is some distance between the means of each class. However, their ability is limited when the classes are too close The AjRC and AAjRC models also performed almost as well as the AdRC classifier demonstrating the potential of using affine policies to update the classifying surface.

6.3. Future directions

In the future, my goal is to extend both decision dependent and connected uncertainty further. First, the unit commitment models used to illustrate uncertain influence are of limited size. My goal is to allow for large models by improving the cut generation procedures. A second important area of research is to allow the use of continuous decision variables as a part of decision-dependent uncertainty sets. In the context of connected uncertainty, a major area for future research is improving the computability of the problem for adaptive optimization problems

in application specific contexts. This is relevant for both robust and distributionally robust optimization. Finally, it is necessary to improve the performance of the adaptive classifier as currently its capabilities are limited. Overall, both of these concepts extend themselves naturally to the study of machine learning and decision-making algorithms in adversarial settings.

References

[1] IEEE 118 power flow test case. https://icseg.iti.illinois.edu/ieee-118-bus-system/. Accessed: 2019.

[2] National climate date center. URL https://www.ncdc.noaa.gov.

[3] Agostinho Agra, Marcio Costa Santos, Dritan Nace, and Michael Poss. A dynamic programming approach for a class of robust optimization problems. *SIAM Journal on Optimization*, 26(3):1799–1823, 2016.

[4] Layth C Alwan, Minghui Xu, Dong-Qing Yao, and Xiaohang Yue. The dynamic newsvendor model with correlated demand. *Decision Sciences*, 47(1):11–30, 2016.

[5] Christoforos Anagnostopoulos, Dimitris K Tasoulis, Niall M Adams, Nicos G Pavlidis, and David J Hand. Online linear and quadratic discriminant analysis with adaptive forgetting for streaming classification. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 5(2):139–166, 2012.

[6] Marcus Ang, Yun Fong Lim, and Melvyn Sim. Robust storage assignment in unit-load warehouses. *Management Science*, 58(11):2114–2130, 2012.

[7] Igor Averbakh and Vasilij Lebedev. Interval data minmax regret network optimization problems. *Discrete Applied Mathematics*, 138(3):289–301, 2004.

[8] Ronald J Balvers and Douglas W Mitchell. Autocorrelated returns and optimal intertemporal portfolio choice. *Management Science*, 43(11):1537–1551, 1997.

[9] Chaithanya Bandi, Eojin Han, and Omid Nohadani. Sustainable inventory with robust periodic-affine policies and application to medical supply chains. *Management Science,*

forthcoming, arXiv preprint arXiv:1806.06744, 2018.

[10] Aharon Ben-Tal and Arkadi Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13, 1999.

[11] Aharon Ben-Tal, Alexander Goryashko, Elana Guslitzer, and Arkadi Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2): 351–376, 2004.

[12] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization*. Princeton University Press, Princeton, NJ, 2009.

[13] Aharon Ben-Tal, Dick Den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

[14] Aharon Ben-Tal, Dick Den Hertog, and Jean-Philippe Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1-2):265–299, 2015.

[15] Dimitri P Bertsekas. *Convex optimization theory*. Athena Scientific Belmont, 2009.

[16] Dimitris Bertsimas and Constantine Caramanis. Finite adaptability in multistage linear optimization. *IEEE Transactions on Automatic Control*, 55(12):2751–2766, 2010.

[17] Dimitris Bertsimas and Melvyn Sim. Robust discrete optimization and network flows. *Mathematical Programming*, 98(1-3):49–71, 2003.

[18] Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations Research*, 52(1): 35–53, 2004.

[19] Dimitris Bertsimas and John N Tsitsiklis. *Introduction to Linear Optimization*. Athena Sci. Ser. Optim. Neural Comput. 6. Athena Scientific, Belmont, MA, 1997.

[20] Dimitris Bertsimas and Phebe Vayanos. Data-driven learning in dynamic pricing using adaptive optimization. `http://www.optimization-online.org/DB_HTML/2014/10/4595.html`, 2014.

[21] Dimitris Bertsimas, Omid Nohadani, and Kwong Meng Teo. Robust optimization for unconstrained simulation-based problems. *Operations Research*, 58(1):161–178, 2010.

[22] Dimitris Bertsimas, David B Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[23] Dimitris Bertsimas, Vishal Gupta, and Nathan Kallus. Robust sample average approximation. *Mathematical Programming*, 171(1):217–282, 2018.

[24] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Stochastic multi-armed-bandit problem with non-stationary rewards. In *Advances in neural information processing systems*, pages 199–207, 2014.

[25] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations research*, 63(5):1227–1244, 2015.

[26] John R Birge and Francois Louveaux. *Introduction to stochastic programming*. Springer Science & Business Media, 2011.

[27] Tim Bollerslev, Ray Y Chou, and Kenneth F Kroner. ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, 52(1-2):5–59, 1992.

[28] Leo Breiman. *Classification and regression trees*. Routledge, 2017.

[29] John M Charnes and W David Kelton. Multivariate autoregressive techniques for constructing confidence regions on the mean vector. *Management Science*, 39(9):1112–1129, 1993.

[30] Millie Chu, Yuriy Zinchenko, Shane G Henderson, and Michael B Sharpe. Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty. *Physics in Medicine & Biology*, 50(23):5463–5477, 2005.

[31] Stephen A Cook. The complexity of theorem-proving procedures. in STOC '71: Proceedings of the Third Annual ACM Symposium on Theory of Computing, ACM, New York, 1971, pages 151–158, 1971.

[32] Kelly J Cormican, David P Morton, and R Kevin Wood. Stochastic network interdiction. *Operations Research*, 46(2):184–197, 1998.

[33] Simon Cousins and John Shawe-Taylor. High-probability minimax probability machines. *Machine Learning*, 106(6):863–886, 2017.

[34] Anderson Rodrigo De Queiroz and David P Morton. Sharing cuts under aggregated forecasts when decomposing multi-stage stochastic programs. *Operations Research Letters*, 41 (3):311–316, 2013.

[35] Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[36] Sarah Jane Delany, Padraig Cunningham, and Alexey Tsymbal. A comparison of ensemble and case-base maintenance techniques for handling concept drift in spam filtering. In *FLAIRS Conference*, pages 340–345, 2006.

[37] Edsger W Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959.

[38] Daniel Duque and David P Morton. Distributionally robust stochastic dual dynamic programming. 2019.

[39] Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166, 2018.

[40] Farzan Farnia and David Tse. A minimax approach to supervised learning. In *Advances in Neural Information Processing Systems*, pages 4240–4248, 2016.

[41] Arnaud Fréville. The multidimensional 0–1 knapsack problem: An overview. *European Journal of Operational Research*, 155(1):1–21, 2004.

[42] Qi Fu, Chee-Khian Sim, and Chung-Piaw Teo. Profit sharing agreements in decentralized supply chains: a distributionally robust approach. *Operations Research*, 66(2):500–513, 2018.

[43] Rui Gao and Anton J Kleywegt. Distributionally robust stochastic optimization with wasserstein distance. *arXiv preprint arXiv:1604.02199*, 2016.

[44] Angelos Georghiou, Wolfram Wiesemann, and Daniel Kuhn. Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming*, 152 (1-2):301–338, 2015.

[45] Laurent El Ghaoui, Maksim Oks, and Francois Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.

[46] Vikas Goel and Ignacio E Grossmann. A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves. *Computers & Chemical Engineering*, 28(8):1409–1429, 2004.

[47] Vikas Goel and Ignacio E Grossmann. A class of stochastic programs with decision dependent uncertainty. *Mathematical Programming*, 108(2-3):355–394, 2006.

[48] Joel Goh and Melvyn Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58(4-part-1):902–917, 2010.

[49] Bin Gu, Xingming Sun, and Victor S Sheng. Structural minimax probability machine. *IEEE Transactions on Neural Networks and Learning Systems*, 28(7):1646–1656, 2016.

[50] Grani A Hanasusanto and Daniel Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Operations Research*, 66(3): 849–869, 2018.

[51] Grani A Hanasusanto, Daniel Kuhn, and Wolfram Wiesemann. K-adaptability in two-stage robust binary programming. *Operations Research*, 63(4):877–891, 2015.

[52] Robert V Hogg and JW McKean. *Introduction to Mathematical Statistics*. Prentice Hall, 2005.

[53] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.

[54] Kjetil Høyland and Stein W Wallace. Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307, 2001.

[55] Dan Andrei Iancu. *Adaptive Robust Optimization with Applications in Inventory and Revenue Management*. Ph.D. thesis, MIT, Cambridge, MA, 2010.

[56] Gerd Infanger and David P Morton. Cut sharing for multistage stochastic linear programs with interstage dependency. *Mathematical Programming*, 75(2):241–256, 1996.

[57] Ruiwei Jiang, Jianhui Wang, and Yongpei Guan. Robust unit commitment with wind power and pumped storage hydro. *IEEE Transactions on Power Systems*, 27(2):800–810, 2012.

[58] Tore W Jonsbråten, Roger J B Wets, and David L Woodruff. A class of stochastic programs with decision dependent random elements. *Annals of Operations Research*, 82:83–106, 1998.

[59] N Bora Keskin and Assaf Zeevi. Chasing demand: Learning and earning in a changing environment. *Mathematics of Operations Research*, 42(2):277–307, 2017.

[60] Scott Kolodziej, Pedro M Castro, and Ignacio E Grossmann. Global optimization of bilinear programs with a multiparametric disaggregation technique. *Journal of Global Optimization*, 57(4):1039–1063, 2013.

[61] Bartosz Krawczyk and Michał Woźniak. One-class classifiers with incremental learning and forgetting for data streams with concept drift. *Soft Computing*, 19(12):3387–3400, 2015.

[62] Daniel Kuhn, Wolfram Wiesemann, and Angelos Georghiou. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, 130(1): 177–209, 2011.

[63] Atsutoshi Kumagai and Tomoharu Iwata. Learning future classifiers without additional data. In *AAAI*, pages 1772–1778, 2016.

[64] Atsutoshi Kumagai and Tomoharu Iwata. Learning non-linear dynamics of decision boundaries for maintaining classification performance. In *AAAI*, pages 2117–2123, 2017.

[65] Atsutoshi Kumagai and Tomoharu Iwata. Learning dynamics of decision boundaries without additional labeled data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1627–1636. ACM, 2018.

[66] Henry Lam. Robust sensitivity analysis for stochastic systems. *Mathematics of Operations Research*, 41(4):1248–1275, 2016.

[67] Gert RG Lanckriet, Laurent El Ghaoui, Chiranjib Bhattacharyya, and Michael I Jordan. A robust minimax approach to classification. *Journal of Machine Learning Research*, 3 (Dec):555–582, 2002.

[68] Chad R Larson, Danko Turcic, and Fuqiang Zhang. An empirical investigation of dynamic ordering policies. *Management Science*, 61(9):2118–2138, 2015.

[69] Hu Li, Ye Wang, Hua Wang, and Bin Zhou. Multi-window based ensemble learning for classification of imbalanced streaming data. *World Wide Web*, 20(6):1507–1525, 2017.

[70] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[71] Miron Livny, Benjamin Melamed, and Athanassios K Tsiolis. The impact of autocorrelation on queuing systems. *Management Science*, 39(3):322–339, 1993.

[72] Alvaro Lorca and Xu Andy Sun. Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind. *IEEE Transactions on Power Systems*, 30(4):1702–1713, 2015.

[73] Alvaro Lorca and Xu Andy Sun. Multistage robust unit commitment with dynamic uncertainty sets and energy storage. *IEEE Transactions on Power Systems*, 32(3):1678–1688, 2017.

[74] Alvaro Lorca, X Andy Sun, Eugene Litvinov, and Tongxin Zheng. Multistage adaptive robust optimization for the unit commitment problem. *Operations Research*, 64(1):32–51, 2016.

[75] Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):645–657, 2017.

[76] Hamed Mamani, Shima Nassiri, and Michael R Wagner. Closed-form solutions for robust inventory management. *Management Science*, 63(5):1625–1643, 2016.

[77] Michele Monaci and Ulrich Pferschy. On the robust knapsack problem. *SIAM Journal on Optimization*, 23(4):1956–1982, 2013.

[78] Roberto Montemanni and Luca Maria Gambardella. The robust shortest path problem with interval data via Benders decomposition. *4OR*, 3(4):315–328, 2005.

[79] Karthik Natarajan, Melvyn Sim, and Joline Uichanco. Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 20(4):695–731, 2010.

[80] Hai-Long Nguyen, Yew-Kwong Woon, and Wee-Keong Ng. A survey on data stream clustering and classification. *Knowledge and information systems*, 45(3):535–569, 2015.

[81] Omid Nohadani and Arkajyoti Roy. Robust optimization with time-dependent uncertainty in radiation therapy. *IISE Transactions on Healthcare Systems Engineering*, 7(2):81–92,

2017.

[82] Omid Nohadani and Kartikey Sharma. Optimization under decision-dependent uncertainty. *SIAM Journal on Optimization*, 28(2):1773–1795, 2018.

[83] Luis J Novoa, Ahmad I Jarrah, and David P Morton. Flow balancing with uncertain demand for automated package sorting centers. *Transportation Science*, 52, 2016.

[84] Michael Poss. Robust combinatorial optimization with variable budgeted uncertainty. *4OR*, 11(1):75–92, 2013.

[85] Michael Poss. Robust combinatorial optimization with variable cost uncertainty. *European Journal of Operational Research*, 237(3):836–845, 2014.

[86] Waseem Rawat and Zenghui Wang. Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9):2352–2449, 2017.

[87] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. MOS-SIAM Ser. Optim. 9, SIAM, Philadelphia, PA, 2009.

[88] Takayuki Shiina and John R Birge. Stochastic unit commitment problem. *International Transactions in Operational Research*, 11(1):19–32, 2004.

[89] Oğuz Solyalı, Jean-François Cordeau, and Gilbert Laporte. The impact of modeling on robust inventory management under demand uncertainty. *Management Science*, 62(4): 1188–1201, 2015.

[90] Allen L Soyster. Technical non-convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157, 1973.

[91] Simon A Spacey, Wolfram Wiesemann, Daniel Kuhn, and Wayne Luk. Robust software partitioning with multiple instantiation. *INFORMS Journal on Computing*, 24(3):500–515, 2012.

[92] Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer Science & Business Media, 2008.

[93] Samer Takriti, John R Birge, and Erik Long. A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems*, 11(3):1497–1508, 1996.

[94] Ruey S Tsay. *Multivariate Time Series Analysis: with R and financial applications*. John Wiley & Sons, 2013.

[95] Canan Uçkun, Audun Botterud, and John R Birge. An improved stochastic unit commitment formulation to accommodate wind uncertainty. *IEEE Transactions on Power Systems*, 31(4):2507–2517, 2015.

[96] Robin Vujanic, Paul Goulart, and Manfred Morari. Robust optimization of schedules affected by uncertain events. *Journal of Optimization Theory and Applications*, 171:1033–1054, 2016.

[97] Gang Wang, Jianshan Sun, Jian Ma, Kaiquan Xu, and Jibao Gu. Sentiment classification: The contribution of ensemble learning. *Decision support systems*, 57:77–93, 2014.

[98] Joseph Whittaker, C Whitehead, and M Somers. A dynamic scorecard for monitoring baseline performance with application to tracking a mortgage portfolio. *Journal of the Operational Research Society*, 58(7):911–921, 2007.

[99] Wolfram Wiesemann, Daniel Kuhn, and Melvyn Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

[100] Lei Wu, Mohammad Shahidehpour, and Tao Li. Stochastic security-constrained unit commitment. *IEEE Transactions on Power Systems*, 22(2):800–811, 2007.

[101] Linwei Xin and David A Goldberg. Distributionally robust inventory control when demand is a martingale. *arXiv preprint arXiv:1511.09437*, 2015.

[102] Shuliang Xu and Junhong Wang. A fast incremental extreme learning machine algorithm for data streams classification. *Expert Systems with Applications*, 65:332–344, 2016.

[103] Gang Yu and Jian Yang. On the robust shortest path problem. *Computers & Operations Research*, 25(6):457–468, 1998.

[104] X. Zhang, M. Kamgarpour, A. Georghiou, P.J. Goulart, and J. Lygeros. Robust optimal control with adjustable uncertainty sets. *Automatica*, 75:249–259, 2017.

[105] Long Zhao and Bo Zeng. Robust unit commitment problem with demand response and wind energy. In *2012 IEEE power and energy society general meeting*, pages 1–8. IEEE, 2012.

[106] Paweł Zieliński. The computational complexity of the relative robust shortest path problem with interval data. *European Journal of Operational Research*, 158(3):570–576, 2004.

[107] Indrė Žliobaitė, Albert Bifet, Bernhard Pfahringer, and Geoffrey Holmes. Active learning with drifting streaming data. *IEEE transactions on neural networks and learning systems*, 25(1):27–39, 2013.

APPENDIX A

A.1. Proof of Theorem 4.2.3

The following is the proof for the robust counterpart for the polyhedral uncertainty set.

Proof. For (P), we define the joint uncertainty set $\mathcal{U}$ as

$$\begin{aligned}\mathcal{U} &= \left\{(\mathbf{d}_1^\top, \mathbf{d}_2^\top, \ldots, \mathbf{d}_T^\top)^\top \mid \mathbf{d}_1 \in \mathcal{U}_1,\ \mathbf{d}_t \in \mathcal{U}_t(\mathbf{d}_{t-1})\ \forall t = 2, \ldots, T\right\} \\ &= \left\{(\mathbf{d}_1^\top, \mathbf{d}_2^\top, \ldots, \mathbf{d}_T^\top)^\top \mid \mathbf{G}_1\mathbf{d}_1 \geq \mathbf{g}_1,\ \mathbf{G}_t\mathbf{d}_t \geq \mathbf{g}_t + \mathbf{\Delta}_t\mathbf{d}_{t-1}\ \forall t = 2, \ldots, T\right\},\end{aligned}$$

combining the polyhedral sets for each period. The robust counterpart of (C-RO) becomes

$$\sum_{t=1}^{T} \mathbf{d}_t^\top \mathbf{x}_t \leq B \ \ \forall (\mathbf{d}_1^\top, \mathbf{d}_2^\top, \ldots, \mathbf{d}_T^\top)^\top \in \mathcal{U},$$

which can be rewritten as

$$\max_{(\mathbf{d}_1^\top, \mathbf{d}_2^\top, \ldots, \mathbf{d}_T^\top)^\top \in \mathcal{U}} \sum_{t=1}^{T} \mathbf{d}_t^\top \mathbf{x}_t \leq B.$$

The LHS is computed by

$$\begin{aligned}\max_{\mathbf{d}_t} \quad & \sum_{t=1}^{T} \mathbf{d}_t^\top \mathbf{x}_t \\ \text{s.t.} \quad & \mathbf{G}_1\mathbf{d}_1 \geq \mathbf{g}_1 \\ & \mathbf{G}_t\mathbf{d}_t \geq \mathbf{g}_t + \mathbf{\Delta}_t\mathbf{d}_{t-1} \qquad \forall t = 2, \ldots, T.\end{aligned}$$

Using duality, the robust counterpart of (C-RO) is

$$\sum_{t=1}^{T} \mathbf{q}_t^\top \mathbf{b}_t \leq B$$

$$\mathbf{q}_t^\top \mathbf{A}_t - \mathbf{q}_{t+1}^\top \boldsymbol{\Delta}_{t+1} = \mathbf{x}_t^\top, \quad \forall t = 1, \ldots, T$$

$$\mathbf{q}_t \leq 0 \quad \forall t = 1, \ldots, T,$$

which is the desired result. □

A.2. ARO with Ellipsoidal CU Sets

In this section, we reformulate the connected constraints from problem (CU-ARO) for ellipsoidal uncertainty sets, where the center depends on the previous period realization as

$$\mathcal{U}_t(\mathbf{d}_{t-1}) = \{\mathbf{d}_t \mid \mathbf{d}_t = \boldsymbol{\mu}_t(\mathbf{d}_{t-1}) + \mathbf{L}_t \mathbf{u}_t : \|\mathbf{u}_t\|_2 \leq r_t\}, \tag{A.1}$$

where $\mathbf{L}_t \mathbf{L}_t^\top = \boldsymbol{\Sigma}_t$. The dependence of $\boldsymbol{\mu}_t$ on the previous period realization is given by

$$\boldsymbol{\mu}_t(\mathbf{d}_{t-1}) = \mathbf{A}_t \boldsymbol{\mu}_{t-1}(\mathbf{d}_{t-2}) + \mathbf{F}_t \mathbf{d}_{t-1} + \mathbf{c}_t. \tag{A.2}$$

For this setting, the following theorem provides the reformulation for CU sets with affine decision policies.

Theorem A.2.1. *The two-period adjustable optimization problem* (CU-ARO) *has a tractable reformulation, when the uncertainty affects the center and the fully adaptive decisions are replaced by affine decision rules.*

Proof. We replace $\mathbf{x}_2(\mathbf{d}_1)$ with the affine decision rule $\mathbf{x}_2(\mathbf{d}_1) = \mathbf{X}_2 \mathbf{d}_1$ and expand $\mathbf{b}_1(\mathbf{d}_1) = \mathbf{B}_1 \mathbf{d}_1$ and $\mathbf{b}_2(\mathbf{d}_2) = \mathbf{B}_2 \mathbf{d}_2$. We focus on reformulating the second constraint, which is affected

by the connected uncertainty and whose robust problem is

$$\begin{aligned}
\max_{\mathbf{d}_1,\mathbf{d}_2}\ & \mathbf{B}_{2,i}^\top \mathbf{d}_2 - [\mathbf{A}_{22}\mathbf{X}_2]_i^\top \mathbf{d}_1 \\
\text{s.t.}\ \ & \mathbf{d}_1 = \boldsymbol{\mu}_1 + \mathbf{L}_1\mathbf{u}_1 \\
& \mathbf{d}_2 = \boldsymbol{\mu}_2(\mathbf{d}_1) + \mathbf{L}_2\mathbf{u}_2 \\
& \|\mathbf{u}_t\|_2 \le r_t \quad \forall t = 1, 2.
\end{aligned}$$

Substituting $\mathbf{d}_1$ and $\mathbf{d}_2$, we rewrite the problem as

$$\begin{aligned}
\max_{\mathbf{d}_1,\mathbf{d}_2}\ & \mathbf{B}_{2,i}^\top(\mathbf{A}_2\boldsymbol{\mu}_1 + \mathbf{F}_2\mathbf{d}_1 + \mathbf{c}_2) + \mathbf{B}_{2,i}^\top\mathbf{L}_2\mathbf{u}_2 - [\mathbf{A}_{22}\mathbf{X}_2]_i^\top\boldsymbol{\mu}_1 - [\mathbf{A}_{22}\mathbf{X}_2]_i^\top\mathbf{L}_1\mathbf{u}_1 \\
\text{s.t.}\ \ & \|\mathbf{u}_t\|_2 \le r_t \quad \forall t = 1, 2.
\end{aligned}$$

Using the dual, the complete second constraint of (CU-ARO) is given by

$$\begin{aligned}
\mathbf{A}_{21,i}^\top\mathbf{x}_1 \ge\ & \mathbf{B}_{2,i}^\top(\mathbf{A}_2\boldsymbol{\mu}_1 + \mathbf{F}_2\boldsymbol{\mu}_1 + \mathbf{c}_2) - [\mathbf{A}_{22}\mathbf{X}_2]_i^\top\boldsymbol{\mu}_1 + r_2\|\mathbf{B}_{2,i}^\top\mathbf{L}_2\|_2 + r_1\|\mathbf{B}_{2,i}^\top\mathbf{F}_2\mathbf{L}_1 \\
& - [\mathbf{A}_{22}\mathbf{X}_2]_i^\top\mathbf{L}_1\|_2.
\end{aligned}$$

The remaining constraints in (CU-ARO) is then reformulated in a similar manner, leading to a tractable reformulation. □

A.3. Robust Counterpart of (C-DRO)

The following proposition better illustrates the connection between constraint (C-DRO) and its robust counterpart (4.5). Given the uncertainty sets $\widetilde{\mathcal{U}}_1, \ldots, \widetilde{\mathcal{U}}_T^{T-1}$, their joint uncertainty set is defined by

$$\widetilde{\mathcal{U}} = \left\{ P \mid P = P_1 \times \cdots \times P_{T|T-1},\ P_1 \in \widetilde{\mathcal{U}}_1,\ P_{t|t-1} \in \widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})\ \forall \mathbf{d}_{t-1} \in \Xi_{t-1}\ \forall t \right\}.$$

The joint set, $\widetilde{\mathcal{U}}$ is the set of all distributions P with the marginals lying in the specified uncertainty sets $\widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})$.

Proposition A.3.1. *Given the sets* $\widetilde{\mathcal{U}}_1, \ldots, \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})$ *and their joint uncertainty set* $\widetilde{\mathcal{U}}$*, the robust counterpart of constraint* (C-DRO)*, given by*

$$\sup_{P \in \widetilde{\mathcal{U}}} \mathbb{E}_P \left[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t) \right] \leq B \tag{A.3}$$

is equivalent to

$$\sup_{P_1 \in \widetilde{\mathcal{U}}_1} \mathbb{E}_{P_1} \Bigg[h_1(\mathbf{x}_1, \mathbf{d}_1) + \sup_{P_{2|1} \in \widetilde{\mathcal{U}}_2(\mathbf{d}_1)} \Bigg\{ \mathbb{E}_{P_{2|1}} \Bigg[h_2(\mathbf{x}_2, \mathbf{d}_2) + \ldots + \sup_{P_{T|T-1} \in \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})} \Big\{ \mathbb{E}_{P_{T|T-1}} \left[h_T(\mathbf{x}_T, \mathbf{d}_T) \right] \Big\} \Bigg] \Bigg\} \Bigg] \leq B.$$

Proof. We first show the forward direction of the equivalence. Observe that for any distribution $P \in \widetilde{\mathcal{U}}$ with the corresponding marginal distribution P_1, and for each $\mathbf{d}_{t-1}$ with the conditional distribution $P_{t|t-1}$, it holds that

$$\begin{aligned} &\mathbb{E}_{P_1} \Big[h_1(\mathbf{x}_1, \mathbf{d}_1) + \Big\{ \mathbb{E}_{P_{2|1}} \Big[h_2(\mathbf{x}_2, \mathbf{d}_2) + \cdots + \Big\{ \mathbb{E}_{P_{T|T-1}} \left[h_T(\mathbf{x}_T, \mathbf{d}_T) \right] \Big\} \Big] \Big\} \Big] \\ &= \mathbb{E}_P \left[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t) \right]. \end{aligned} \tag{A.4}$$

For a small $\epsilon > 0$, let $P^* \in \widetilde{\mathcal{U}}$ be such that

$$\mathbb{E}_{P^*} \left[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t) \right] \geq \sup_{P \in \widetilde{\mathcal{U}}} \mathbb{E}_P \left[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t) \right] - \epsilon. \tag{A.5}$$

That means the LHS of (A.5) is ϵ-optimal to the LHS of (A.3). Since P^* is in $\widetilde{\mathcal{U}}$, there exist marginal distributions $P_1^*, \ldots, P_{T|T-1}^*$ such that P_1^* lies in $\widetilde{\mathcal{U}}_1$ and the conditional distribution of $\mathbf{d}_t$, $P_{t|t-1}^*$ lies in $\widetilde{\mathcal{U}}(\mathbf{d}_{t-1})$, where $\mathbf{d}_{t-1} \in \Xi_{t-1}$. This holds true for all $t = 2, \ldots, T$. Using (A.4),

this means

$$\sup_{P_1\in\widetilde{\mathcal{U}}_1}\mathbb{E}_{P_1}\Big[h_1(\mathbf{x}_1,\mathbf{d}_1)+\sup_{P_{2|1}\in\widetilde{\mathcal{U}}_2(\mathbf{d}_1)}\Big\{\mathbb{E}_{P_{2|1}}\Big[h_2(\mathbf{x}_2,\mathbf{d}_2)+\ldots \\ +\sup_{P_{T|T-1}\in\widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})}\Big\{\mathbb{E}_{P_{T|T-1}}\left[h_T(\mathbf{x}_T,\mathbf{d}_T)\right]\Big\}\Big]\Big\}\Big]$$

$$\geq \mathbb{E}_{P_1^*}\left[h_1(\mathbf{x}_1,\mathbf{d}_1)+\left\{\mathbb{E}_{P^*_{2|1}}\left[h_2(\mathbf{x}_2,\mathbf{d}_2)+\cdots+\left\{\mathbb{E}_{P^*_{T|T-1}}\left[h_T(\mathbf{x}_T,\mathbf{d}_T)\right]\right\}\right]\right\}\right]$$

$$=\mathbb{E}_{P^*}[\sum_{t=1}^{T}h_t(\mathbf{x}_t,\mathbf{d}_t)]$$

$$\geq \sup_{P\in\widetilde{\mathcal{U}}}\mathbb{E}_P[\sum_{t=1}^{T}h_t(\mathbf{x}_t,\mathbf{d}_t)]-\epsilon.$$

Now for the opposite side of the inequality, let $P_1^*, P_{2|1}^*$ up to $P_{T|T-1}^*$ be ϵ-optimal to (4.5). The constraints of (4.5) ensure that $P_1^*\in\widetilde{\mathcal{U}}_1$ and $P_{t|t-1}^*\in\widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})$. Since $\widetilde{\mathcal{U}}$ is the set of all joint distributions with these specified marginals, for the joint distribution P^* with these marginals, the equation (A.4) will hold. However since P_1^* and $P_{t|t-1}^*$ are ϵ-optimal, this means that

$$\sup_{P_1\in\widetilde{\mathcal{U}}_1}\mathbb{E}_{P_1}\Big[h_1(\mathbf{x}_1,\mathbf{d}_1)+\sup_{P_{2|1}\in\widetilde{\mathcal{U}}_2(\mathbf{d}_1)}\Big\{\mathbb{E}_{P_{2|1}}\Big[h_2(\mathbf{x}_2,\mathbf{d}_2)+\ldots \\ +\sup_{P_{T|T-1}\in\widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})}\Big\{\mathbb{E}_{P_{T|T-1}}\left[h_T(\mathbf{x}_T,\mathbf{d}_T)\right]\Big\}\Big]\Big\}\Big]$$

$$\leq \mathbb{E}_{P_1^*}\Big[h_1(\mathbf{x}_1,\mathbf{d}_1)+\Big\{\mathbb{E}_{P^*_{2|1}}\Big[h_2(\mathbf{x}_2,\mathbf{d}_2)+\cdots+\Big\{\mathbb{E}_{P^*_{T|T-1}}\left[h_T(\mathbf{x}_T,\mathbf{d}_T)\right]\Big\}\Big]\Big\}\Big]+\epsilon$$

$$=\mathbb{E}_{P^*}\left[\sum_{t=1}^{T}h_t(\mathbf{x}_t,\mathbf{d}_t)\right]+\epsilon$$

$$\leq \sup_{P\in\widetilde{\mathcal{U}}}\mathbb{E}_P[\sum_{t=1}^{T}h_t(\mathbf{x}_t,\mathbf{d}_t)]+\epsilon.$$

This gives the opposite inequality, and the result follows by letting ϵ shrink to zero. □

Note that this result does not depend on the structure of the moment based uncertainty set and can be extended to sets $\widetilde{\mathcal{U}}_1, \ldots, \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})$ of any structure, in which the parameters depend on previous realizations. An important part of this result is the additive nature of the constraint.

A.4. RO Application: Knapsack Problem with Negative Correlations

We repeat the experiments conducted for the Knapsack problem with a negative correlation among the weights over time i.e., $\mathbf{\Psi} = -0.2 \times \mathbf{I}$. Figure A.1 presents the value of average objective for any level of constraint satisfaction (left) and the number of non-zero allocations vs. set size. Figure A.2 displays how the average constraint satisfaction (left) and the average objective value (right) change with size of the uncertainty set r. These results show that when uncertainties are negatively correlated, CU solutions achieve a better objective value but a lower constraint satisfaction than NC solutions. However, for any level of constraint satisfaction both have similar performance.

- ***Effect of Uncertainty Set Size***: For both models, constraint satisfaction increases and average objective value decreases with r. Note that the objective value is only measured, if constraints are satisfied.
- ***CU vs. NC***: For any r, CU solutions have lower constraint satisfaction than NC solutions. CU solutions also have higher average objective value. This is because connected sets depend on the first period, and the negative correlation instead of magnifying the worst-case causes one of the realizations to take a lower value. This reduces the protection but increases the objective value.

  Note that for any level of satisfaction, the average objective of CU is almost the same as NC (see Figure A.2 left).

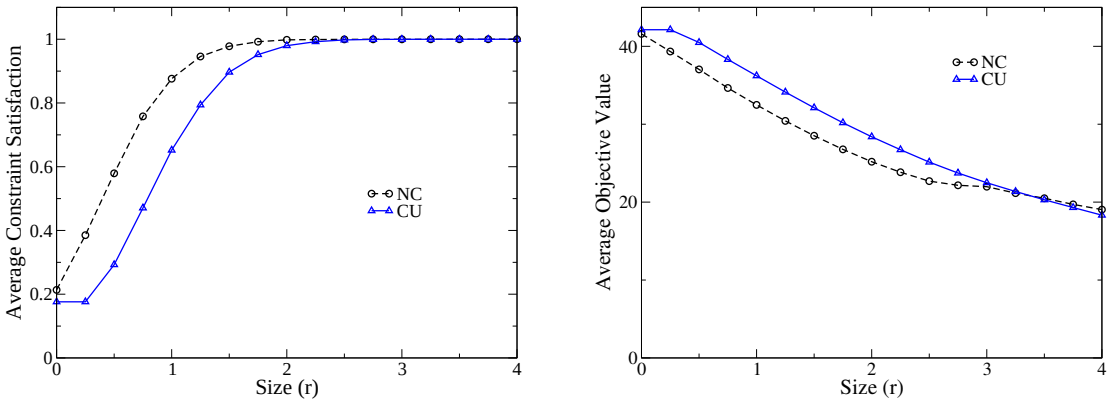


Figure A.1. Comparison of connected and non-connected sets for the robust knapsack problem at different set sizes: (left) average fraction of constraint satisfaction, and (right) average objective value.

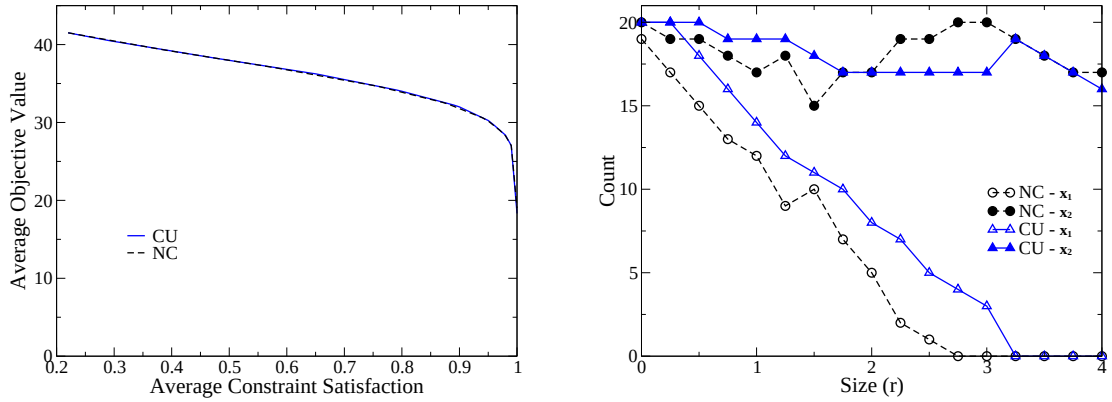


Figure A.2. Comparison of connected and non-connected sets: (left) average objective value vs. constraint satisfaction, and (right) for a single iteration, the number of non-zero variables of a period.

- ***First vs Second period solutions***: For a single estimate in (ii) shown in Figure A.2 (right), the optimal solution gradually concentrates only on $\mathbf{x}_2$ for both CU and NC as r increases. For NC, this is because $\mathbf{c}_2$ tends to be higher. For CU, the negative correlation prevents the magnification of the weights for both periods as such second period allocations which have higher objective coefficients can be selected.

A.5. DRO Application: Portfolio Optimization

We study a practical portfolio optimization problem on historical stock data. For our portfolio, we choose among 5 stocks. The experiment is repeated 150 times for randomly selected dates. At each date, we compute the weekly returns for the previous 100 weeks based on stock price data and fit them to a time series model. The two weeks following the selected dates serve to evaluate the performance of the model.

To capture risk aversion, in each experiment, we maximize a concave piecewise linear utility function $u = \min(1.5r, 0.015 + r, 0.06 + 0.1r)$, where r denotes the return on the portfolio. It is assumed that $\overline{\boldsymbol{\mu}}_2 = \boldsymbol{\mu}_2(\mathbf{d}_1) + \boldsymbol{\delta}$ and $\underline{\boldsymbol{\mu}}_2 = \boldsymbol{\mu}_2(\mathbf{d}_1) - \boldsymbol{\delta}$ with $\boldsymbol{\mu}_2(\mathbf{d}_1) = \boldsymbol{\mu}_0 + \mathbf{A}\boldsymbol{\mu}_1 + \mathbf{B}\mathbf{d}_1$. The vector $\boldsymbol{\mu}_0$ and the matrices $\mathbf{A}$ and $\mathbf{B}$ are estimated from data via a vector autoregressive moving average with lag 1 [94]. The parameters $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are set to the residual covariance matrices. The value of $\boldsymbol{\mu}_1$ is the return point-estimate at the end of the first week and $\boldsymbol{\delta}$ is three times the standard error of this estimate (in order to cover almost all realized means under normality assumption).

To probe the performance of the proposed CU model, we compare it against the standard DRO and a modified DRO, described as following:

- **:** CU: Model with connected uncertainty set,
- **:** DRO-1: Model with $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1$, and
- **:** DRO-2: Model with $\boldsymbol{\mu}_2 = \mathbf{A}\boldsymbol{\mu}_1 + \mathbf{B}\boldsymbol{\mu}_1$.

DRO-1 represents the standard model as used widely in the literature. It is computationally attractive because of a simpler reformulation. However, it does not take into account potential connections to previous periods. We propose a modified version (DRO-2), which captures some of the potential connections to previous periods using the unconditional mean for the second period.

In these experiments, the parameters $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are defined based on past weeks, and the returns are computed over the future two weeks. We evaluate the solution quality from these three models by comparing the returns. Each experiment starts with an initial wealth of $W_0 = \$100$, which is recomputed using the total return on the portfolio

$$W_{t+1} = W_t \cdot (\widehat{\mathbf{d}}_t^\top \mathbf{x}_t),\ t = 1, 2,$$

The realized demand $\widehat{\mathbf{d}}_t$ is taken from the actual stock data. This wealth (W) is then averaged over the different experiments and reported in Table A.1, along with the standard deviation (Std) as a measure for robustness of the solution.

Model	Period 1		Period 2		Period 3	
	W	Std	W	Std	W	Std
CU	100	0	100.1	1.97	101.2	5.38
DRO1	100	0	99.9	2.36	100.9	5.92
DRO2	100	0	99.9	2.36	100.9	6.06

Table A.1. Average wealth (W) and its standard deviation (Std) over time for various models

Table A.1 displays the performances of the three different models. For all three, we observe a positive average return at the end of two weeks, which is attributed to the mean positive return on the random samples for the five stocks. Furthermore, the standard deviation (spread) of the sample paths grows as time elapses, because the inherent uncertainty in returns is compounding over time.

When comparing the three models, we observe marginal differences in the average wealth, reflecting the relatively short time horizon. However, we observe the wealth standard deviation for the CU model to be lower than both of the DRO models. This is attributed to the fact that the CU model captures the compounding worst-case effect of connected periods and yields a more conservative solution. the CU model reduces the wealth standard deviation for wealth

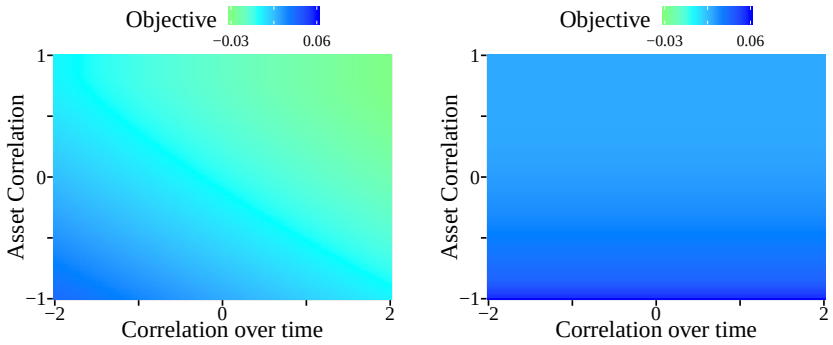


Figure A.3. CU and DRO problem objective

by 0.417 (17.6% of the original wealth standard deviation) for period 1 and 0.563 (9.5%) for period 2 compared to DRO-1. Other measures of deviation, such as interquartile range, reveal a similar decrease in wealth. Therefore, the CU model is able to select assets that are less volatile in order to provide more robust allocations.

Note that solving the CU model is computationally more demanding than either of the DRO models, because of latter's convex subproblems. Therefore, depending on the application, the advantages of the CU model have to outweigh the additional computational burden. Furthermore, the DRO-2 model represents another way to leverage the connection between uncertainties of different periods. However, it only accounts for the average effect between periods and not the worst case.

A.6. DRO Application: Portfolio Optimization II

(1) *Objective function:* to measure performance of the model in the nominal setting (Figure A.3).

(2) *Worst-case realized wealth:* to study the effect of adversarial settings (Figure A.5).

(3) *Standard deviation of wealth:* to evaluate the robustness of solutions (Figure A.4).

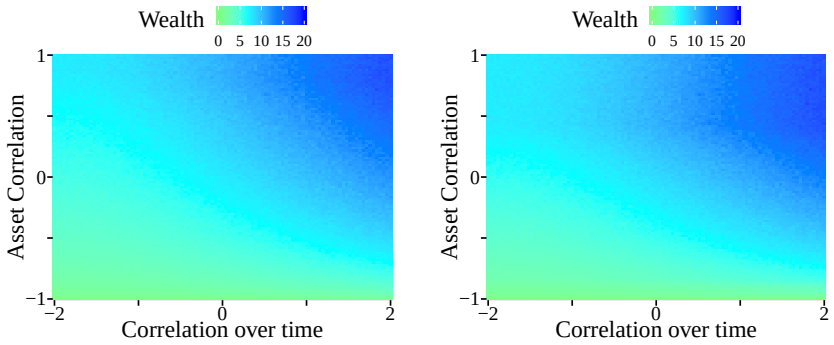


Figure A.4. Realized wealth standard deviation for CU and DRO at end of period 2

Objective

Asset Correlation. Figure A.3 indicates that positive correlations among the assets reduce the objective performance for the CU model. This is due to the fact that positive correlation worsens the worst-case returns everywhere simultaneously, whereas a negative correlation prevents this from occurring as a result the objective (which is worst-case by definition) reduces.

Correlation over time. Figure A.3 also presents similar behavior for correlation over time leading to worse objective performance for the same reasons as asset correlation. The DRO model does not capture the correlation over time but it also performs better when the assets are negatively correlated.

Standard Deviation and Worst-Case Wealth

Figure A.4 shows the standard deviation for the realized wealth at the end of the second period.

Asset Correlation. The standard deviation is higher, when the assets are positively correlated among themselves and over time.

Figure A.5 shows the worst-case realized wealth at the end of the second period. The worst case wealth is higher, when assets are negatively correlated among themselves and over time. This is because the negative correlation makes it unlikely for both returns to be very low. The

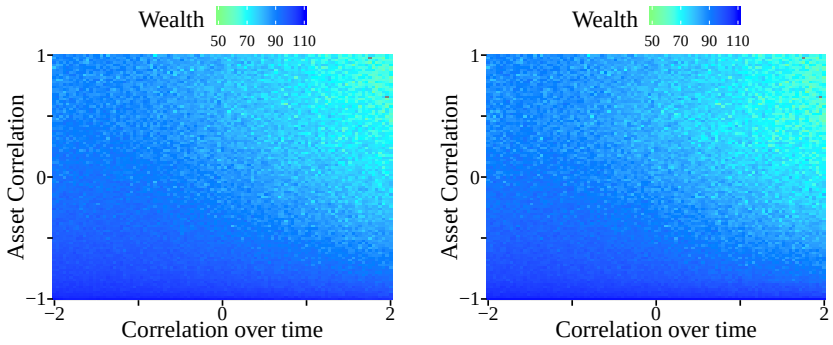


Figure A.5. Worst-case realized wealth for period 2

pattern observed here is similar to that in Figure A.3, with the worst-case wealth being worse when assets are positively correlated.

A.7. Proof of Theorem 5.4.1

Proof. We can rewrite problem (AjRC) as

$$\begin{aligned}
&\min_{r,\overline{\mathbf{U}},\overline{\mathbf{V}}} \; r \\
&\text{s.t. } (\overline{\mathbf{X}} + \mathbf{L}_x\overline{\mathbf{U}} - \overline{\mathbf{Y}} - \mathbf{L}_y\overline{\mathbf{W}})\boldsymbol{\xi} = \mathbf{0} && \forall \boldsymbol{\xi} \in \mathcal{U} \\
&\quad \|\overline{\mathbf{U}}\boldsymbol{\xi}\|_2 \leq r && \forall \boldsymbol{\xi} \in \mathcal{U} \\
&\quad \|\overline{\mathbf{W}}\boldsymbol{\xi}\|_2 \leq r && \forall \boldsymbol{\xi} \in \mathcal{U}.
\end{aligned}$$

Consider the constraint $\|\overline{\mathbf{U}}\boldsymbol{\xi}\|_2 \leq r \;\; \forall \boldsymbol{\xi} \in \mathcal{U}$. Expanding $\overline{\mathbf{U}}\boldsymbol{\xi} = \mathbf{u}_0 + \mathbf{U}_x\mathbf{x} + \mathbf{U}_y\mathbf{y}$ the constraint becomes

$$\|\mathbf{u}_0 + \mathbf{U}_x\mathbf{x} + \mathbf{U}_y\mathbf{y}\|_2 \leq r \quad \forall \mathbf{x} \in \mathcal{U}_x, \; \mathbf{y} \in \mathcal{U}_y,$$

where $\mathcal{U}_x, \mathcal{U}_y$ are ellipsoids centered around $\boldsymbol{\mu}_x^0, \boldsymbol{\mu}_y^0$ with covariance matrices $\boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_y$ and radius κ_x, κ_y. Substituting the same we get the following constraint

$$\|\mathbf{u}_0 + \mathbf{U}_x\boldsymbol{\mu}_x^0 + \mathbf{U}_y\boldsymbol{\mu}_y^0 + \mathbf{U}_x\mathbf{L}_x\mathbf{u} + \mathbf{U}_y\mathbf{L}_y\mathbf{w}\|_2 \leq r \quad \forall (\mathbf{u} \in \|\mathbf{u}\|_2 \leq \kappa_x), (\mathbf{w} \in \|\mathbf{w}\|_2 \leq \kappa_y).$$

Using the Lemma A.7.1, we can reformulate the above constraint as a semi-definite inequality

$$\begin{bmatrix} r - \lambda_{xx}\kappa_x^2 - \lambda_{xy}\kappa_y^2 & \mathbf{u}_0^\top + (\boldsymbol{\mu}_x^0)^\top \mathbf{U}_x^\top + (\boldsymbol{\mu}_y^0)^\top \mathbf{U}_y^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{u}_0 + \mathbf{U}_x\boldsymbol{\mu}_x^0 + \mathbf{U}_y\boldsymbol{\mu}_y^0 & r\mathbf{I} & \mathbf{U}_x\mathbf{L}_x & \mathbf{U}_y\mathbf{L}_y \\ \mathbf{0} & \mathbf{L}_x^\top\mathbf{U}_x^\top & \lambda_{xx}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_y^\top\mathbf{U}_y^\top & \mathbf{0} & \lambda_{xy}\mathbf{I} \end{bmatrix} \succeq 0$$

$$\lambda_{xx}, \lambda_{xy} \geq 0.$$

Repeating a similar reformulation for the constraint $\|\overline{\mathbf{W}}\boldsymbol{\xi}\|_2 \leq r \;\; \forall \boldsymbol{\xi} \in \mathcal{U}$, we can write the problem (AjRC) as

$$\min_{r, \overline{\mathbf{U}}, \overline{\mathbf{V}}} \; r$$

$$\text{s.t. } (\overline{\mathbf{X}} + \mathbf{L}_x\overline{\mathbf{U}} - \overline{\mathbf{Y}} - \mathbf{L}_y\overline{\mathbf{W}})\boldsymbol{\xi} = \mathbf{0} \qquad \forall \boldsymbol{\xi} \in \mathcal{U}$$

$$\begin{bmatrix} r - \lambda_{xx}\kappa_x^2 - \lambda_{xy}\kappa_y^2 & \mathbf{u}_0^\top + (\boldsymbol{\mu}_x^0)^\top \mathbf{U}_x^\top + (\boldsymbol{\mu}_y^0)^\top \mathbf{U}_y^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{u}_0 + \mathbf{U}_x\boldsymbol{\mu}_x^0 + \mathbf{U}_y\boldsymbol{\mu}_y^0 & r\mathbf{I} & \mathbf{U}_x\mathbf{L}_x & \mathbf{U}_y\mathbf{L}_y \\ \mathbf{0} & \mathbf{L}_x^\top\mathbf{U}_x^\top & \lambda_{xx}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_y^\top\mathbf{U}_y^\top & \mathbf{0} & \lambda_{xy}\mathbf{I} \end{bmatrix} \succeq 0$$

$$\begin{bmatrix} r - \lambda_{yx}\kappa_x^2 - \lambda_{yy}\kappa_y^2 & \mathbf{w}_0^\top + (\boldsymbol{\mu}_x^0)^\top \mathbf{W}_x^\top + (\boldsymbol{\mu}_y^0)^\top \mathbf{W}_y^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{w}_0 + \mathbf{W}_x\boldsymbol{\mu}_x^0 + \mathbf{W}_y\boldsymbol{\mu}_y^0 & r\mathbf{I} & \mathbf{W}_x\mathbf{L}_x & \mathbf{W}_y\mathbf{L}_y \\ \mathbf{0} & \mathbf{L}_x^\top\mathbf{W}_x^\top & \lambda_{yx}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_y^\top\mathbf{W}_y^\top & \mathbf{0} & \lambda_{yy}\mathbf{I} \end{bmatrix} \succeq 0$$

$$\lambda_{xx}, \lambda_{xy}, \lambda_{yx}, \lambda_{yy} \geq 0.$$

Here, we have $\overline{\mathbf{U}} = [\mathbf{u}_0 \mid \mathbf{U}_x \mid \mathbf{U}_y]$ and $\overline{\mathbf{W}} = [\mathbf{w}_0 \mid \mathbf{W}_x \mid \mathbf{W}_y]$. Next, if we assume that the uncertainty set $\mathcal{U} = 1 \times \mathcal{U}_x \times \mathcal{U}_y$ is full dimensional, i.e., any element in the real space with the same dimension as $\mathcal{U}$ can be written as a linear combinations of elements in $\mathcal{U}$, then we can write the constraint $(\overline{\mathbf{X}} + \mathbf{L}_x\overline{\mathbf{U}} - \overline{\mathbf{Y}} - \mathbf{L}_y\overline{\mathbf{W}})\boldsymbol{\xi} = \mathbf{0} \;\; \forall \boldsymbol{\xi} \in \mathcal{U}$ as

$$\overline{\mathbf{X}} + \mathbf{L}_x\overline{\mathbf{U}} - \overline{\mathbf{Y}} - \mathbf{L}_y\overline{\mathbf{W}} = \mathbf{0}.$$

This is equivalent to

$$[\mathbf{a}_x \mid \mathbf{A}_{xx} \mid \mathbf{A}_{xy}] + \mathbf{L}_x[\mathbf{u}_0 \mid \mathbf{U}_x \mid \mathbf{U}_y] = [\mathbf{a}_y \mid \mathbf{A}_{yx} \mid \mathbf{A}_{yy}] + \mathbf{L}_y[\mathbf{w}_0 \mid \mathbf{W}_x \mid \mathbf{W}_y].$$

This simplifies to the following set of constraints

$$\begin{aligned}
\mathbf{a}_x + \mathbf{L}_x\mathbf{u}_0 &= \mathbf{a}_y + \mathbf{L}_y\mathbf{w}_0 \\
\mathbf{A}_{xx} + \mathbf{L}_x\mathbf{U}_x &= \mathbf{A}_{yx} + \mathbf{L}_y\mathbf{W}_x \\
\mathbf{A}_{xy} + \mathbf{L}_x\mathbf{U}_y &= \mathbf{A}_{yy} + \mathbf{L}_y\mathbf{W}_y.
\end{aligned}$$

Therefore, the optimization problem can be expressed as

$$\min_{r,\overline{\mathbf{U}},\overline{\mathbf{V}}} \; r$$

$$\text{s.t. } \mathbf{a}_x + \mathbf{L}_x\mathbf{u}_0 = \mathbf{a}_y + \mathbf{L}_y\mathbf{w}_0$$

$$\mathbf{A}_{xx} + \mathbf{L}_x\mathbf{U}_x = \mathbf{A}_{yx} + \mathbf{L}_y\mathbf{W}_x$$

$$\mathbf{A}_{xy} + \mathbf{L}_x\mathbf{U}_y = \mathbf{A}_{yy} + \mathbf{L}_y\mathbf{W}_y$$

$$\begin{bmatrix} r - \lambda_{xx}\kappa_x^2 - \lambda_{xy}\kappa_y^2 & \mathbf{u}_0^\top + (\boldsymbol{\mu}_x^0)^\top\mathbf{U}_x^\top + (\boldsymbol{\mu}_y^0)^\top\mathbf{U}_y^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{u}_0 + \mathbf{U}_x\boldsymbol{\mu}_x^0 + \mathbf{U}_y\boldsymbol{\mu}_y^0 & r\mathbf{I} & \mathbf{U}_x\mathbf{L}_x & \mathbf{U}_y\mathbf{L}_y \\ \mathbf{0} & \mathbf{L}_x^\top\mathbf{U}_x^\top & \lambda_{xx}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_y^\top\mathbf{U}_y^\top & \mathbf{0} & \lambda_{xy}\mathbf{I} \end{bmatrix} \succeq 0$$

$$\begin{bmatrix} r - \lambda_{yx}\kappa_x^2 - \lambda_{yy}\kappa_y^2 & \mathbf{w}_0^\top + (\boldsymbol{\mu}_x^0)^\top\mathbf{W}_y^\top + (\boldsymbol{\mu}_y^0)^\top\mathbf{W}_y^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{w}_0 + \mathbf{W}_x\boldsymbol{\mu}_x^0 + \mathbf{W}_y\boldsymbol{\mu}_y^0 & r\mathbf{I} & \mathbf{W}_x\mathbf{L}_x & \mathbf{W}_y\mathbf{L}_y \\ \mathbf{0} & \mathbf{L}_x^\top\mathbf{W}_x^\top & \lambda_{yx}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_y^\top\mathbf{W}_y^\top & \mathbf{0} & \lambda_{yy}\mathbf{I} \end{bmatrix} \succeq 0$$

$$\lambda_{xx}, \lambda_{xy}, \lambda_{yx}, \lambda_{yy} \geq 0.$$

□

Lemma A.7.1. *The following two constraints are equivalent*

(i) $\|\mathbf{a} + \sum_{t=1}^{T} \mathbf{A}_t\mathbf{u}_t\|_2 \leq r \;\; \forall(\mathbf{u}_1, \ldots \mathbf{u}_T : \|\mathbf{u}_t\|_2 \leq \kappa_t, \; \forall t).$

(ii) There exist $\lambda_1, \ldots, \lambda_T \geq 0$ *such that*

$$\begin{bmatrix} r - \sum_{t=1}^{T} \lambda_t \kappa_t^2 & \mathbf{a}^\top & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{a} & r\mathbf{I} & \mathbf{A}_1 & \mathbf{A}_2 & \ldots & \mathbf{A}_T \\ \mathbf{0} & \mathbf{A}_1^\top & \lambda_1 \mathbf{I} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^\top & \mathbf{0} & \lambda_2 \mathbf{I} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ \mathbf{0} & \mathbf{A}_T^\top & \mathbf{0} & \mathbf{0} & & \lambda_T \mathbf{I} \end{bmatrix} \succeq 0. \tag{A.6}$$

Proof. The constraint $\|\mathbf{a} + \sum_{t=1}^{T} \mathbf{A}_t \mathbf{u}_t\|_2 \leq r \Leftrightarrow \|\mathbf{a} + [\mathbf{A}_1 \ \mathbf{A}_2 \ \ldots \mathbf{A}_T] \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_T \end{bmatrix} \|_2 \leq r$ can be written as the following semi-definite constraint

$$\begin{bmatrix} r & \mathbf{a}^\top + \sum_{t=1}^{T} \mathbf{u}_t^\top \mathbf{A}_t^\top \\ \mathbf{a} + \sum_{t=1}^{T} \mathbf{A}_t \mathbf{u}_t & r\mathbf{I} \end{bmatrix} \succeq 0.$$

Thus the robust version of the constraint can be written as

$$\begin{bmatrix} r & \mathbf{a}^\top + \sum_{t=1}^{T} \mathbf{u}_t^\top \mathbf{A}_t^\top \\ \mathbf{a} + \sum_{t=1}^{T} \mathbf{A}_t \mathbf{u}_t & r\mathbf{I} \end{bmatrix} \succeq 0 \ \ \forall \|\mathbf{u}_t\|_2 \leq \kappa_t \ \forall t = 1, \ldots, T.$$

This can be expressed as a quadratic using a vector $[p, \mathbf{q}^\top]^\top$, where $p \in \mathbb{R}$ and $\mathbf{q} \in \mathbb{R}^n$ via

$$rp^2 + 2p\mathbf{q}^\top(\mathbf{a} + \sum_{t=1}^{T} \mathbf{A}_t \mathbf{u}_t) + r\mathbf{q}^\top \mathbf{q} \geq 0 \ \ \forall (\mathbf{u}_1, \ldots, \mathbf{u}_T : \|\mathbf{u}_t\|_2 \leq \kappa_t \ \forall t = 1, \ldots, T).$$

Expanding the terms, we can write the above as

$$rp^2 + 2p\mathbf{q}^\top \mathbf{a} + \sum_{t=1}^{T} 2p\mathbf{q}^\top \mathbf{A}_t \mathbf{u}_t + r\mathbf{q}^\top \mathbf{q} \geq 0 \ \ \forall (\mathbf{u}_1, \ldots, \mathbf{u}_T : \|\mathbf{u}_t\|_2 \leq \kappa_t \ \forall t = 1, \ldots, T),$$

which is equivalent to

$$rp^2 + 2p\mathbf{q}^\top \mathbf{a} + r\mathbf{q}^\top \mathbf{q} + \sum_{t=1}^{T} \min_{\mathbf{u}_t : \|\mathbf{u}_t\|_2 \le \kappa_t} 2p\mathbf{q}^\top \mathbf{A}_t \mathbf{u}_t \ge 0, \ \ \forall [p, \mathbf{q}]$$

$$\iff rp^2 + 2p\mathbf{q}^\top \mathbf{a} + r\mathbf{q}^\top \mathbf{q} - \sum_{t=1}^{T} 2p\kappa_t \|\mathbf{q}^\top \mathbf{A}_t\|_2 \ge 0, \ \ \forall [p, \mathbf{q}]$$

$$\iff rp^2 + 2p\mathbf{q}^\top \mathbf{a} + r\mathbf{q}^\top \mathbf{q} + \sum_{t=1}^{T} 2\mathbf{q}^\top \mathbf{A}_t \boldsymbol{\xi}_t \ge 0, \ \ \forall [p, \mathbf{q}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T] : \boldsymbol{\xi}_t^\top \boldsymbol{\xi}_t \le p^2 \kappa_t^2, \ \forall t = 1, \ldots, T.$$

Next, we express the above constraints in terms of matrices. First, the quadratic equation on the left hand side can be expressed as

$$(p, \mathbf{q}^\top, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi_2}^\top, \ldots, \boldsymbol{\xi}_T^\top) \begin{bmatrix} r & \mathbf{a}^\top & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{a} & r\mathbf{I} & \mathbf{A}_1 & \mathbf{A}_2 & \ldots & \mathbf{A}_T \\ \mathbf{0} & \mathbf{A}_1^\top & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^\top & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ \mathbf{0} & \mathbf{A}_T^\top & \mathbf{0} & \mathbf{0} & & \mathbf{0} \end{bmatrix} \begin{pmatrix} p \\ \mathbf{q} \\ \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \\ \vdots \\ \boldsymbol{\xi}_T \end{pmatrix} \ge 0. \tag{A.7}$$

For the set $\left\{(p, \mathbf{q}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_T) : \boldsymbol{\xi}_1^\top \boldsymbol{\xi}_1 \le p^2 \kappa_1^2\right\}$, the above can be rewritten as

$$(p, \mathbf{q}^\top, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi_2}^\top, \ldots, \boldsymbol{\xi}_T^\top) \begin{bmatrix} \kappa_1^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & & \mathbf{0} \end{bmatrix} \begin{pmatrix} p \\ \mathbf{q} \\ \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \\ \vdots \\ \boldsymbol{\xi}_T \end{pmatrix} \ge 0. \tag{A.8}$$

This can be repeated for all such sets. Using the S-lemma, the constraint (A.7) holds for all $(p, \mathbf{q}^\top, \boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top, \ldots, \boldsymbol{\xi}_T^\top)$ that satisfy constraint (A.8), if and only if there exists a $\lambda_1 \geq 0$ such that

$$\begin{bmatrix} r - \lambda_1 \kappa_1^2 & \mathbf{a}^\top & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{a} & r\mathbf{I} & \mathbf{A}_1 & \mathbf{A}_2 & \ldots & \mathbf{A}_T \\ \mathbf{0} & \mathbf{A}_1^\top & \lambda_1 \mathbf{I} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^\top & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ \mathbf{0} & \mathbf{A}_T^\top & \mathbf{0} & \mathbf{0} & & \mathbf{0} \end{bmatrix} \succeq 0. \tag{A.9}$$

Repeating this until $t = T$, we obtain the following constraint

$$\begin{bmatrix} r - \sum_{t=1}^T \lambda_t \kappa_t^2 & \mathbf{a}^\top & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{a} & r\mathbf{I} & \mathbf{A}_1 & \mathbf{A}_2 & \ldots & \mathbf{A}_T \\ \mathbf{0} & \mathbf{A}_1^\top & \lambda_1 \mathbf{I} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^\top & \mathbf{0} & \lambda_2 \mathbf{I} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ \mathbf{0} & \mathbf{A}_T^\top & \mathbf{0} & \mathbf{0} & & \lambda_T \mathbf{I} \end{bmatrix} \succeq 0, \tag{A.10}$$

which concludes the proof. □