Tractable Policies in Dynamic Robust Optimization

Omar El Housni

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020

Abstract

Tractable Policies in Dynamic Robust Optimization

Omar El Housni

In many sequential decision problems, uncertainty is revealed over time and we need to make decisions in the face of uncertainty. This is a fundamental problem arising in many applications such as facility location, resource allocation and capacity planning under demand uncertainty. Robust optimization is an approach to model uncertainty where we optimize over the worst-case realization of parameters within an uncertainty set. While computing an optimal solution in dynamic robust optimization is usually intractable, affine policies (or linear decision rules) are widely used as an approximate solution approach. However, there is a stark contrast between the observed good empirical performance and the bad worst-case theoretical performance bounds. In the first part of this thesis, we address this stark contrast between theory and practice. In particular, we introduce a probabilistic approach in Chapter 2 to analyze the performance of affine policies on randomly generated instances and show they are near-optimal with high probability under reasonable assumptions. In Chapter 3, we study these policies under important models of uncertainty such as budget of uncertainty sets and intersection of budgeted sets and show that affine policies give an optimal approximation matching the hardness of approximation. In the second part of the thesis and based on our analysis of affine policies, we design new tractable policies for dynamic robust optimization. In particular, in Chapter 4, we present a tractable framework to design piecewise affine policies that can be computed efficiently and improve over affine policies for many instances. In Chapter 5, we introduce extended affine policies and threshold policies and show that their performance guarantees are significantly better than previous policies. Finally, in Chapter 6, we study piecewise static policies and their limitations for solving some classes of dynamic robust optimization problems.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First, and foremost, I would like to express my deep gratitude to my advisor Vineet Goyal. My PhD achievements would not have been possible without him. He has shaped me into the researcher I am today and has provided me with unconditional support and valuable guidance since my first day at Columbia. Vineet: I will never forget the countless hours we spent together working on challenging research problems. It has been a great joy learning from you and I am extremely grateful for your commitment to both my academic and personal development. Thank you for believing in me and for guiding my steps in the academic life and beyond.

I would also like to thank my thesis committee members and collaborators: Garud Iyengar, Donald Goldfarb, David Shmoys, Clifford Stein and Aharon Ben-Tal. I am very fortunate and honored to have such amazing researchers in my committee. Garud: thank you for your critical feedback on my research and for your support and help throughout my journey at Columbia. Your advices were very valuable especially at crucial moments of my PhD. Don: thank you for teaching me linear and convex optimization. Your core class on optimization has significantly nurtured my early interests in the field. I will always remember the many long and enjoyable conversations we had in the past five years. David: I was very fortunate to work with you in the last two years of my PhD. I have learned a lot from you on combinatorial optimization, but more importantly, you have been a great mentor to me during my job market year and I am very grateful and indebted to you. Cliff: thank you for teaching me so many interesting classes on Algorithms and Optimization. I really enjoyed working with you on matching problems and I look forward to continue our collaboration. Aharon, your research on robust optimization has always inspired my work and particularly this thesis. I had the pleasure to collaborate with you early in my PhD and I am glad to see the result of our collaboration in Chapter 4 of this thesis .

I had the pleasure to interact with a wonderful group of researchers at Columbia and learned from numerous wonderful professors. I would like to thank all the IEOR and DRO faculty who taught me so many exciting courses and helped me gain perspective on my research path. Special thanks go to Omar Besbes, Soulaymane Kachani and Adam Elmachtoub for being great mentors and for supporting me throughout all the steps of my PhD. I would also like to thank all the IEOR administrative staff who were of great help during my time at Columbia and made IEOR a social and vibrant enjoyable environment, especially Lizbeth, Kristen and Jaya. In addition, thanks to Amine Bennouna who worked with me on some parts of Chapter 5 while visiting IEOR as a research intern.

The IEOR community has been a wonderful home where I made many great friends. The journey would not have been as delightful without their presence. I would like to thank all my fellow PhD students for the amazing moments we've spent together, from taking classes, to working on research projects, playing sports, enjoying NYC, traveling and having fun. Thank you: Oussama, Kumar, Raghav, Antoine, Amine Allouah, Jalaj, Julien, Nouri, Achraf, Gonzalo, Chun, Omar Mouchtaki, Mike, Vashist, Apurv, Irene, Vlada, Suraj, Min, Agathe, Enrique, Rajan...

Last, but not least, I would like to thank my family for always supporting me. I thank my little brother and sister for their constant cheer and encouragement. Finally, from the bottom of my heart, I would like to thank my parents for their unconditional love and unwavering support. I owe them more than I could ever express in words. To them I dedicate this thesis.

To my parents Ahmed and Souad

Introduction

In most sequential decision problems, uncertainty is revealed over time and we need to make decisions in the face of uncertainty. This is a fundamental problem arising in almost every business application where real-time decisions are based on the information revealed thus far. For instance, in capacity planning problems, retailers need to make capacity decisions while the uncertain demand is sequentially revealed in the market. In facility location problems, manufacturers need to decide the location of the stores before they observe the uncertain demand requests from customers.

Stochastic and robust optimization are two widely used paradigms to handle uncertainty. In the stochastic optimization approach, uncertainty is modeled as a probability distribution and the goal is to optimize an expected objective [1]. We refer the reader to Kall and Wallace [2], Prekopa [3], Shapiro [4], Shapiro et al. [5], Birge and Louveaux [6] for a detailed discussion on stochastic optimization. On the other hand, in the robust optimization approach, we consider an adversarial model of uncertainty using an uncertainty set and the goal is to optimize over the worst-case realization from the uncertainty set. This approach was first introduced by Soyster [7] and has been extensively studied in recent years. While robust optimization approach might seem conservative, the decision maker can control the level of conservatism by choosing an appropriate uncertainty set. Moreover, designing an uncertainty set from historical data is significantly less challenging than estimating a joint probability distribution especially for high-dimensional uncertainty. Furthermore, robust optimization leads to a tractable approach where a feasible static solution can be computed efficiently for a large class of problems. However, computing an optimal (or dynamic) solution can be hard in general in both the stochastic and robust paradigms due to the "curse of dimensionality". This intractability of computing the optimal adjustable solution necessitates considering approximate solution policies. We refer the reader to Ben-Tal and Nemirovski [8, 9, 10], El Ghaoui and Lebret [11], Bertsimas and Sim [12, 13], Goldfarb and Iyengar [14], Bertsimas et al. [15] and Ben-Tal et al. [16] for a detailed discussion of robust optimization.

In this thesis, we focus on the robust optimization framework to model uncertainty. Our goal is to design and analyze tractable approximation policies and algorithms for dynamic robust optimization problems that have both provable theoretical guarantees and can be implemented efficiently in practice. In addition to practical implementation, the worst case performance analysis allows us to understand both the power and limitations of the approximate policies and provides insights towards designing more general policies.

Commonly used approximations policies in robust optimization include functional policies such as static and affine policies where the decision in any period t is restricted to a static or a linear function of the sample path until period t. Both static and affine policies have been studied extensively in the literature and can be computed efficiently for a large class of problems. While the worst-case performance of such approximate policies can be significantly bad as compared to the optimal dynamic solution, the empirical performance, especially of affine policies, has been observed to be near-optimal in a broad range of computational experiments. In the first part of this thesis (Chapters 2 and 3), we aim to bridge the gap between the theoretical and empirical performance of affine policies by providing an extensive theoretical analysis of their performance for a wide range of dynamic robust problems. While affine policies provide good theoretical and empirical approximation in many settings, their performance could be bad for some classes of uncertainty sets. This motivates us to consider more general policies namely piecewise policies where we divide the uncertainty set into several pieces and specify an affine or a static solution for each piece. A significant challenge in designing a practical piecewise policy is to construct good pieces of the uncertainty set. Based on the insights in our analysis of affine policies, we develop new piecewise policies that improve significantly over affine and static policies in many settings.

In particular, we present a tractable framework to design different classes of piecewise policies and analyze their performance for a fairly general class of robust optimization problems. We discuss piecewise affine policies in Chapter 4, threshold and extended affine policies in Chapter 5 and piecewise static policies in Chapter 6.

This thesis is organized as follows. Chapter 1 is an introduction chapter where we present an overview of the robust optimization problems that we discuss in this thesis. In particular, we present the framework of two-stage adjustable robust optimization. We introduce both the two-stage robust problems with covering constraints and with packing constraints and review couple of preliminaries and known results in the literature. Note that most of the chapters in this thesis would focus on covering problems. We include in Chapter 1 an extensive summary of all our contributions in this thesis. In Chapters 2 and 3, we address the stark contrast between the worst-case theoretical performance and near-optimal empirical performance of affine policies. In particular, we present a probabilistic analysis of affine policies in Chapter 2 that provides a theoretical justification of the good empirical performance of affine policies on random instances of a fairly general class of robust optimization problems. In Chapter 3, we provide a theoretical study on the performance of affine policy on realistic instances under a widely used class of uncertainty sets including budget of uncertainty sets and intersection of budgeted sets. In Chapter 4, we present a tractable framework to design piecewise affine policies that can be computed efficiently and improve over affine policies for a wide range of instances. In Chapter 5, we introduce extended affine policies and threshold policies and show that they improve significantly over the previous known policies for many instances. Finally, in Chapter 6, we study piecewise static policies and their limitations for solving some classes of dynamic robust optimization problems.

Chapter 1: Two-stage robust optimization

In this thesis, we focus on a fairly general class of two-stage robust optimization problems, also known as two-stage adjustable robust optimization, that arises in many applications. The dynamics of this class of problems are such that the decision maker takes a first stage decision before observing the realization of uncertain parameters. Then, adversary selects the uncertain parameters from an uncertainty set. Finally, the decision maker takes a recourse decision after observing the realization of these uncertain parameters. In this chapter, we first introduce the class of two-stage robust problems with covering linear constraints and review related literature. Then, we present a summary of all our contributions in Chapters 2, 3, 4, 5 that would mainly focus on this class of problems. At the end of this chapter, we introduce the class of two-stage robust problems with packing linear constraints that we discuss in Chapter 6 and summarize the contributions of Chapter 6.

1.1 Two-stage robust optimization with covering constraints

Consider the following two-stage adjustable robust linear optimization problem with uncertain right hand side:

$$
\begin{aligned}
z_{\mathsf{AR}}(\boldsymbol{c},\boldsymbol{d},\boldsymbol{A},\boldsymbol{B},\mathcal{U}) = \min_{\boldsymbol{x}}\ & \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \min_{\boldsymbol{y}(\boldsymbol{h})} \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{h}) \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) \ \geq\ \boldsymbol{h}, \quad \forall \boldsymbol{h}\in\mathcal{U} \\
& \boldsymbol{x}\in\mathcal{X} \\
& \boldsymbol{y}(\boldsymbol{h})\in\mathbb{R}^n_+, \quad \forall \boldsymbol{h}\in\mathcal{U},
\end{aligned}
\tag{1.1}
$$

where $\boldsymbol{A}\in\mathbb{R}^{m\times n}, \boldsymbol{B}\in\mathbb{R}^{m\times n}_+, \boldsymbol{c}\in\mathbb{R}^n_+$ and $\boldsymbol{d}\in\mathbb{R}^n_+$. The right-hand-side $\boldsymbol{h}$ is uncertain and belongs to a compact convex uncertainty set $\mathcal{U}\subseteq\mathbb{R}^m_+$. The recourse matrix is non-negative and fixed, i.e.,

$\boldsymbol{B}$ belongs to the non-negative orthant and does not depend on the uncertain parameter $\boldsymbol{h}$. The goal in this problem is to select the first-stage decision $\boldsymbol{x} \in \mathcal{X}$, where $\mathcal{X}$ is a polyhedral set and the second-stage recourse decision, $\boldsymbol{y}(\boldsymbol{h})$, as a function of the uncertain right hand side realization, $\boldsymbol{h}$ such that the worst-case cost over all realizations of $\boldsymbol{h} \in \mathcal{U}$ is minimized.

This model has been widely considered in the literature (see for example Bertsimas and de Ruiter [17], Bertsimas and Goyal [18], Dhamdhere et al. [19], El Housni and Goyal [20], Gupta et al. [21], Xu and Burer [22], Zhen et al. [23].) It captures many important applications including set cover, capacity planning, facility location and network design problems under uncertain demand. Here, the right hand side $\boldsymbol{h}$ models the uncertain demand and the covering constraints capture the requirement of satisfying the uncertain demand. However, the adjustable robust optimization problem (1.1) is intractable in general. In fact, Feige et al. [24] show that the two-stage adjustable problem (1.1) can not be approximated within a ratio better than $\Omega(\frac{\log n}{\log \log n})$ under a reasonable complexity assumption, namely, 3SAT can not be solved in time $2^{O(\sqrt{n})}$ on instances of size n.

In view of the intractability, several approximation policies (or *decision rules*) have been considered in the literature for (1.1) including static, piecewise static, affine and piecewise affine policies. In a static policy, we compute a single optimal solution $(\boldsymbol{x}, \boldsymbol{y})$ that is feasible for all realizations of the uncertain right hand side. Bertsimas et al. [25] relate the performance of static solution to the symmetry of the uncertainty set and show that it provides a good approximation to the adjustable problem if the uncertainty set verifies some symmetry properties. However, static policy is too conservative in general and the performance of static solutions can be arbitrarily large for a general convex uncertainty set.

Ben-Tal et al. [26] introduce affine policy approximation for (1.1), where they restrict the second-stage decision, $\boldsymbol{y}(\boldsymbol{h})$ to being an affine function of the uncertain right-hand-side $\boldsymbol{h}$, i.e., $\boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}$ for some decision variables $\boldsymbol{P} \in \mathbb{R}^{n \times m}$ and $\boldsymbol{q} \in \mathbb{R}^n$. Affine policy can be computed efficiently for a large class of uncertainty sets and therefore, provide a tractable approximation for the two-stage problem. Furthermore, the empirical performance of affine policies has been observed to be near-optimal for a large class of instances even though theoretically, optimality of affine policies is known in only a few settings. Bertsimas et al. [27] and Iancu et al. [28] show that affine policy is optimal for multi-stage adjustable problems with a single uncertain parameter at each stage. Bertsimas and Goyal [18] show that affine policy is optimal for the two-stage adjustable problem (1.1) if the uncertainty set $\mathcal{U}$ is a simplex. However, in the particular case where we assume only non-negativity constraints on the first stage decision variable, i.e. $\boldsymbol{x} \geq \boldsymbol{0}$, they show that the worst-case performance of an optimal affine solution is $\Theta(\sqrt{m})$ times the optimal cost of (1.1) [18]. Note that the gap could be even larger for general polyhedral constraints that involves only $\boldsymbol{x}$ i.e., $\boldsymbol{x} \in \mathcal{X}$. Therefore, there is a significant gap between the worst-case performance of affine policies and the observed empirical performance.

More general decision rules have been considered in the literature for two-stage problems; binary decision rules (Bertsimas and Georghiou [29]), adjustable solutions via iterative splitting of uncertainty sets (Postek and Den Hertog [30]), k-adaptibility (Hanasusanto et al. [31], El Housni and Goyal [32]), segregated linear decision rules (Chen et al. [33]), Fourier–Motzkin elimination (Zhen et al. [34]), etc. While these decision rules can improve in some instances over affine policies, they become computationally very challenging especially for large size instances. For a more extensive review of the literature, we refer the reader to Bertsimas et al. [15], Ben-Tal et al. [16] and Yanikoglu et al. [35]

1.1.1 Affine policies: Preliminaries.

Affine policies (also known as *linear decision rules*) are widely used in the literature of robust optimization. They were introduced by Ben-Tal et al. [26] for the two-stage adjustable problem (1.1). In an affine solution, we restrict the second stage decision $\boldsymbol{y}(\boldsymbol{h})$ to be an affine function of the uncertain parameter $\boldsymbol{h}$, i.e.,

$$\boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{P}\boldsymbol{h} + \boldsymbol{q},$$

and we optimize over the variables $\boldsymbol{P}$ and $\boldsymbol{q}$. The affine problem is formulated as:

$$
\begin{aligned}
z_{\mathsf{Aff}}\left(\boldsymbol{c},\boldsymbol{d},\boldsymbol{A},\boldsymbol{B},\mathcal{U}\right) = \min_{\boldsymbol{x},\boldsymbol{P},\boldsymbol{q}} \; & \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \boldsymbol{d}^T\left(\boldsymbol{P}\boldsymbol{h}+\boldsymbol{q}\right) \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\left(\boldsymbol{P}\boldsymbol{h}+\boldsymbol{q}\right) \;\geq\; \boldsymbol{h}, \quad \forall \boldsymbol{h}\in\mathcal{U} \\
& \boldsymbol{P}\boldsymbol{h}+\boldsymbol{q} \;\geq\; \boldsymbol{0}, \quad \forall \boldsymbol{h}\in\mathcal{U} \\
& \boldsymbol{x}\in\mathcal{X}.
\end{aligned}
\tag{1.2}
$$

Affine policy has been widely used as an approximation to (1.1) due to its tractability. In fact, Ben-Tal et al. [26] show that affine problems have an equivalent standard LP formulation when the uncertainty set is described by a polyhedral set. The size of the LP is polynomial in the size of the input parameters (i.e., number of variables and constraints in (1.1) and number of constraints in $\mathcal{U}$). For completeness, we briefly discuss the tractability and compact LP formulation of affine policies. Consider the following polyhedral uncertainty set

$$
\mathcal{U} = \{\boldsymbol{h}\in\mathbb{R}^m_+ \mid \boldsymbol{R}\boldsymbol{h}\leq\boldsymbol{r}\}, \tag{1.3}
$$

where $\boldsymbol{R}\in\mathbb{R}^{L\times m}$ and $\boldsymbol{r}\in\mathbb{R}^L$. The affine problem (1.2) can be reformulated as the following epigraph formulation:

$$
\begin{aligned}
z_{\mathsf{Aff}}\left(\boldsymbol{c},\boldsymbol{d},\boldsymbol{A},\boldsymbol{B},\mathcal{U}\right) = \min \; & \boldsymbol{c}^T\boldsymbol{x} + z \\
& z \geq \boldsymbol{d}^T\left(\boldsymbol{P}\boldsymbol{h}+\boldsymbol{q}\right), \quad \forall \boldsymbol{h}\in\mathcal{U} \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\left(\boldsymbol{P}\boldsymbol{h}+\boldsymbol{q}\right) \;\geq\; \boldsymbol{h}, \quad \forall \boldsymbol{h}\in\mathcal{U} \\
& \boldsymbol{P}\boldsymbol{h}+\boldsymbol{q} \;\geq\; \boldsymbol{0}, \quad \forall \boldsymbol{h}\in\mathcal{U} \\
& \boldsymbol{x}\in\mathcal{X},\; \boldsymbol{P}\in\mathbb{R}^{n\times m},\; \boldsymbol{q}\in\mathbb{R}^n,\; z\in\mathbb{R}.
\end{aligned}
$$

Note that this formulation can have infinitely many constraints but the separation problem is tractable. For example, the separation problem for the first set of constraints is: Given $z, \boldsymbol{x}, \boldsymbol{P}, \boldsymbol{q}$

decide if

$$z - \boldsymbol{d}^T \boldsymbol{q} \geq \max_{\boldsymbol{h} \geq \boldsymbol{0}} \{\boldsymbol{d}^T \boldsymbol{P} \boldsymbol{h} \mid \boldsymbol{R} \boldsymbol{h} \leq \boldsymbol{r}\}. \tag{1.4}$$

This can be done efficiently by solving the above maximization LP. Moreover, Ben-Tal et al. [26] show that we can formulate (1.2) as a compact LP using standard techniques from duality. For instance, consider the first set of constraints (1.4), by taking the dual of the maximization problem, the constraint becomes

$$z - \boldsymbol{d}^T \boldsymbol{q} \geq \min_{\boldsymbol{v} \geq \boldsymbol{0}} \{\boldsymbol{r}^T \boldsymbol{v} \mid \boldsymbol{R}^T \boldsymbol{v} \geq \boldsymbol{P}^T \boldsymbol{d}\}.$$

We can then drop the min and introduce $\boldsymbol{v}$ as a variable. Hence, we obtain the following linear constraints:

$$z - \boldsymbol{d}^T \boldsymbol{q} \geq \boldsymbol{r}^T \boldsymbol{v}, \qquad \boldsymbol{R}^T \boldsymbol{v} \geq \boldsymbol{P}^T \boldsymbol{d}, \qquad \boldsymbol{v} \geq \boldsymbol{0}.$$

We can apply the same techniques for the other constraints. For completeness, we restate the compact LP formulation of Ben-Tal et al. [26] adapted to our problem in Lemma A.0.2. The lemma and its proof are deferred to Appendix A.

1.1.2 Affine policies: Summary of contributions of Chapters 2 and 3

The goal of Chapters 2 and 3 is to address the stark contrast between the worst-case and empirical performance of affine policies and provide a fine-grained analysis of affine policies beyond worst-case.

Chapter 2: Beyond worst-case: a probabilistic analysis of affine policies. In this chapter, we present a theoretical analysis of the performance of affine policies for synthetic instances of two-stage robust optimization problem generated from a probabilistic model. More specifically, we consider random instances of the two-stage adjustable problem (1.1) where the coefficients of the constraint matrix $\boldsymbol{B}$ are randomly generated and analyze the performance of affine policies for a large class of distributions. The main contributions of this chapter are summarized below.

Random Constraint Coefficients. We consider probabilistic instances of (1.1) where the columns of $\boldsymbol{B}$ are generated from independent multivariate distributions, (i.e., for all $j \in [n]$, column $\boldsymbol{B}_j$

is generated from the multivariate distribution $\mathcal{F}_j$ independent from the other columns) and show that affine policy is provably a good approximation with high probability with a bound that is significantly better than the worst-case bound for a large class of distributions including distributions with bounded support and distributions with gaussian and sub-gaussian tails.

1. **Distributions with Bounded Support**. We first consider the case where the support of distributions $\mathcal{F}_j,\ j \in [n]$ is bounded in, say $[0, b]^m$. For all $i \in [m]$, let $\mathcal{F}_{ij}$ denote the marginal distribution of B_{ij} where the column $\boldsymbol{B}_j$ is distributed according to $\mathcal{F}_j$, and let $\mu_{ij} = \mathbb{E}[B_{ij}]$. We show that for sufficiently large values of m and n, affine policy gives a b/μ-approximation to the adjustable problem (1.1) where

$$\mu = \min_{i \in [m]} \frac{1}{n} \sum_{j=1}^{n} \mu_{ij}.$$

   More specifically, with probability at least $(1 - 1/m)$, we have that

$$z_{\mathsf{Aff}}(\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{A}, \boldsymbol{B}, \mathcal{U}) \leq \frac{b}{\mu(1-\epsilon)} \cdot z_{\mathsf{AR}}(\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{A}, \boldsymbol{B}, \mathcal{U}),$$

   where $\epsilon = b/\mu \sqrt{\log m/n}$ (Theorem 2.2.1 in Chapter 2).

   This bound is significantly better than the worst-case approximation bound of $O(\sqrt{m})$ for many distributions. As an example, consider the special case where all coefficients B_{ij} are i.i.d. according to some distribution with bounded support $[0, b]$ and expectation μ. Then affine policy gives b/μ-approximation to the two-stage adjustable problem (1.1) with high probability. Moreover, if the distribution is *symmetric* (such as uniform or Bernoulli distribution with parameter $p = 1/2$), affine policy gives a 2-approximation for the adjustable problem (1.1).

2. **Distributions with Sub-Gaussian tails**. While the above analysis leads to a good approximation for many distributions, the ratio $\frac{b}{\mu}$ can be significantly large in general; for instance, for distributions where extreme values of the support are extremely rare and significantly

far from the mean. In such instances, the bound b/μ can be quite loose. We can tighten the analysis by using the concentration properties of distributions and can extend the analysis even for the case of unbounded support. In particular, we consider the case where for all $j \in [n]$, column $\boldsymbol{B}_j$ is distributed according to a multivariate distribution, $\mathcal{F}_j$ with (possibly) unbounded support and a sub-gaussian tail independent of other columns. Then for sufficiently large values of m and n, with probability at least $(1 - 1/m)$,

$$z_{\mathsf{Aff}}(\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{A}, \boldsymbol{B}, \mathcal{U}) \leq O(\sqrt{\log m + \log n}) \cdot z_{\mathsf{AR}}(\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{A}, \boldsymbol{B}, \mathcal{U}).$$

Here we assume that the parameters of the distributions are independent of the problem dimension. We prove the case of distributions with sub-gaussian tails in Theorem 2.2.3 of Chapter 2.

We would like to note that the above performance bounds are in stark contrast with the worst case performance bound $O(\sqrt{m})$ for affine policies that is tight. For the random instances where columns of $\boldsymbol{B}$ are independent according to above distributions, with high probability the performance bound is significantly better. Therefore, our results provide a theoretical justification of the good empirical performance of affine policies and close the gap between worst case bound of $O(\sqrt{m})$ and observed empirical performance. Furthermore, surprisingly these performance bounds are independent of the structure of the uncertainty set, $\mathcal{U}$ unlike in previous work where the performance bounds depend on the geometric properties of $\mathcal{U}$. Our analysis is based on a *dual-reformulation* of (1.1) introduced in [17] where (1.1) is reformulated as an alternate two-stage adjustable optimization and the uncertainty set in the alternate formulation depends on the constraint matrix $\boldsymbol{B}$. Using the probabilistic structure of $\boldsymbol{B}$, we show that the alternate *dual* uncertainty set is close to a simplex for which affine policies are optimal.

We would also like to note that our performance bounds are not necessarily tight and the actual performance on particular instances can be even better. We test the empirical performance of affine policies for random instances generated according to uniform and folded normal distributions and observe that affine policies are nearly optimal with a worst optimality gap of 4% (i.e. approximation ratio of 1.04) on our test instances as compared to the optimal adjustable solution that we compute using a mixed integer program (MIP).

Worst-case distribution for Affine policies. While affine policies give with high probability a good approximation for random instances according to a large class of commonly used distributions, we present a distribution where the performance of affine policies is $\Omega(\sqrt{m})$ with high probability for instances generated from this distribution. In particular, there is no smoothed analysis for affine policies. Moreover, this bound matches the worst-case deterministic bound for affine policies. We would like to remark that in the worst-case distribution, the coefficients B_{ij} depend on the dimension of the problem. This suggests that to obtain bad instances for affine policies, we need to generate instances using a structured distribution where the structure of the distribution might depend on the problem structure.

Chapter 3: Affine policies for budget of uncertainty sets. In this chapter, we study the performance of affine policies for *realistic* instances of the two-stage adjustable problem (1.1) (in particular the instances of (1.1) are not drawn randomly from a class of distributions as in the previous chapter). The focus here is to analyze the performance of affine policie for an important class of uncertainty sets widely used in practice, namely budget of uncertainty. Again, one of our important goals in this analysis is to address the stark contrast between the observed near-optimal empirical performance and the worst-case approximation bound of $\Theta(\sqrt{m})$ [18]. Towards this, we consider the class of budget of uncertainty sets and intersection of budget of uncertainty sets that was introduced in Bertsimas and Sim [13]. This is widely used class of uncertainty sets in practice where the decision maker can specify a budget on the sum of adversarial deviations of the uncertain parameter from the nominal values. In particular, a budget of uncertainty set can be formulated as follows:

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} w_i h_i \leq 1 \right\}, \tag{1.5}$$

where $w_i \in [0,1]$ for all $i \in [m]$. It is known that the two-stage adjustable problem (1.1) is hard to approximate under this class of uncertainty set. In particular, Feige et al. [24] show that the two-stage adjustable problem (1.1) where $\mathcal{U}$ is the budget of uncertainty set (1.5) is hard to approximate within a factor $\Omega(\frac{\log n}{\log \log n})$ even when all w_i are equal, $\boldsymbol{A}, \boldsymbol{B}$ are 0-1 matrices and $\mathcal{X} = \mathbb{R}^n_+$. The main contributions of this chapter are the following.

(a) **Optimal approximation for budget of uncertainty sets.** We show that affine policy gives an optimal approximation for the two-stage adjustable robust problem for budgeted uncertainty sets. In particular, affine policy gives an $O(\frac{\log n}{\log \log n})$-approximation to the two-stage adjustable problem (1.1) where $\mathcal{U}$ is a budget of uncertainty set (1.5). This performance bound matches the hardness of approximation [24]; thereby, showing that surprisingly affine policies give an optimal approximation (up to some constant factor) for (1.1) for budget of uncertainty sets. In other words, there is no polynomial time algorithm with worst-case approximation guarantee better than affine policies by more than some constant factor. This bound significantly improves over the previous known bound of $O(\sqrt{m})$ [18, 36] for budget of uncertainty sets. Moreover, our result holds for general polyhedral constraints on the first stage variable $\boldsymbol{x} \in \mathcal{X}$. In particular, we can model for example upper bounds on $\boldsymbol{x}$, this is in contrast with the previous bounds in the literature that have been shown only in the special case of $\mathcal{X} = \mathbb{R}^n_+$.

    Our analysis relies on constructing a feasible affine solution whose worst-case cost is within a factor $O(\frac{\log n}{\log \log n})$ of the optimal cost. In particular, we partition the components of $\mathcal{U}$ into *inexpensive* and *expensive* components based on a threshold and construct an affine solution that covers only the inexpensive components using a linear solution. The remaining components are covered using a static solution. We show that for an appropriately chosen threshold that depends on the optimal cost, such an affine solution gives an $O(\frac{\log n}{\log \log n})$-approximation for the two-stage problem for the budget of uncertainty set.

    Therefore, in addition to establishing the performance bound that matches the hardness of approximation, our analysis shows there is a near-optimal affine solution whose structure is closely related to *threshold policies* that are widely used in many applications. This structural property might be of independent interest and also gives an alternate faster algorithm for computing near-optimal affine solutions for budget of uncertainty sets as we discuss later.

(b) **Intersection of budgeted sets.** We also consider a more general family of uncertainty sets, namely the following intersection of budget of uncertainty sets:

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i \in S_\ell} w_{\ell i} h_i \leq 1 \;\; \forall \ell \in [L] \right\}, \tag{1.6}$$

where $\boldsymbol{w}_l \in [0,1]^m$, and $S_\ell \subseteq [m]$ for all $\ell \in [L]$. The set (1.6) is defined by the intersection of L budget constraints. These are an important generalization of the budget of uncertainty set (1.5) that are widely used in practice. They capture for instance CLT sets [37] and *inclusion-constrained budgeted* sets [38].

(i) We first consider the case when the family of subsets S_ℓ for $\ell \in [L]$ are disjoint. We refer to this class of sets as *disjoint constrained budgeted sets*. These are essentially Cartesian product of L budget of uncertainty sets. We show that affine policy is near-optimal and gives an $O\left(\log^2 n / \log\log n\right)$-approximation to (1.1) for this class of sets. We would like to note that the bound is independent of L. Similar to the case of budget of uncertainty sets, our analysis is based on constructing a near-optimal affine solution by partitioning components of $\mathcal{U}$ into *inexpensive* and *expensive* components using appropriate thresholds for each of the L budgeted sets in the Cartesian product. However, in this case, we are able to relate the performance of our affine solution to only a lower bound of $z_{\mathsf{AR}}(\mathcal{U})$. In particular, we use an *online algorithm* for the *fractional covering problem* to both construct thresholds (and therefore, a feasible affine solution) as well as the lower bound of the optimal value.

(ii) For general intersection of L budgets. Under the assumption that $\mathcal{X}$ is a polyhedral cone (for example $\mathcal{X} = \mathbb{R}^n_+$), we show that affine policy gives $O\left(\log L \log n / \log\log n\right)$ to (1.1) for the case where $\mathcal{U}$ is *permutation invariant*. We say that $\mathcal{U}$ is permutation invariant if for any $\boldsymbol{h} \in \mathcal{U}$ and any τ permutation of $[m]$, then $\boldsymbol{h}^\tau \in \mathcal{U}$ where $h_i^\tau = h_{\tau(i)}$. This class captures many important sets such as CLT sets. The performance of affine policy depends on L in this case but degrades gracefully. For general intersection of budgeted sets and $\mathcal{X}$ a polyhedral cone, we show a worst-case bound of $O\left(L \log n / \log \log n\right)$ on the performance of affine policy for (1.1). We summarize our results in Table 1.1.

(c) **Faster algorithm to compute near-optimal affine solutions.** Based on the structural properties of the near-optimal affine policies constructed for analysis of performance, we present an alternate algorithm to compute an approximate affine policy for (1.1) for budget of uncertainty sets that is significantly faster than computing optimal affine policy by solving a large LP. In particular, our construction partitions the components into *inexpensive* and *expensive* based on a threshold depending on the optimal cost and shows the existence of a near-optimal affine solution that covers a fraction of *inexpensive* components using a linear solution and the remaining components using a static solution.

From an algorithmic perspective, while we do not know the optimal cost and therefore, the threshold, we can still use this structure of a near-optimal affine solution to construct a good affine solution. In particular, our approximate affine solution can be computed efficiently by solving a single LP covering problem with $O(n + m)$ second stage variables as opposed to $O(nm)$ second stage variables in the optimal affine policy. Our algorithm scales very well and is significantly faster than computing affine policies. For instance, for $m = n = 100$, it takes several minutes to compute the optimal affine policy whereas our algorithm computes an approximate affine policy within a few seconds. The comparison becomes even more stark when we increase the problem size. In particular, for $m = n = 200$, the average time for optimal affine policy is more than an hour, whereas our algorithm computes an approximate affine policy in under 2 minutes. Moreover, our solution remains within 15% of the optimal affine solution and the sub-optimality gap does not increase with dimension in our numerical experiments. We would like to note that since our approximate affine is based on the construction of affine policy in our analysis, the worst-case approximation bound for the faster algorithm is also $O(\frac{\log n}{\log\log n})$.

(d) **General constraint matrices.** We show that the assumption on the non-negativity of the recourse matrix $\boldsymbol{B}$ is crucial for obtaining any non-trivial theoretical bounds on the performance of affine policies. We give a family of instances of the two-stage adjustable problem where the recourse matrix $\boldsymbol{B}$ is a network matrix with entries in $\{-1,0,1\}$ and show that the gap between optimal affine and adjustable policies can be unbounded even for the single budget of uncertainty set. The second-stage matrix being a network matrix captures important applications including lot sizing and facility location.

We also show that our results do not extend to the case of uncertainty in the left hand side. In particular, we give a family of instances of two-stage adjustable problem with a first stage matrix $\boldsymbol{A}$ that depends on the uncertain parameter $\boldsymbol{h}$ and show that the gap between optimal affine and adjustable policies can be as bad as $\Omega(\max(m,n))$ even for the special case of box uncertainty sets.

	Uncertainty sets	Our Bounds
1.	Budget of uncertainty set (1.5)	$O\left(\frac{\log n}{\log\log n}\right)^{*}$
2.	Disjoint Intersection of Budgeted sets (3.10)	$O\left(\frac{\log^2 n}{\log\log n}\right)$
3.	Permutation Invariant Intersection of Budgeted sets (1.6)	$O\left(\frac{\log L\cdot\log n}{\log\log n}\right)^{**}$
4.	General Intersection of Budgeted sets (1.6)	$O\left(\frac{L\cdot\log n}{\log\log n}\right)^{**}$

Table 1.1: Our performance bounds for affine policy under different uncertainty sets including budget of uncertainty set and intersection of budgeted sets. * denotes that this bound mathches the approximation hardness of (1.1). ** denotes that these bounds hold under the assumption that $\mathcal{X}$ is a polyhedral cone.

1.1.3 Piecewise policies: Summary of contributions of Chapters 4 and 5

Chapters 4 and 5 focus on the design and analysis of new policies that improve significantly over affine policies including piecewise affine policies (Chapter 4), threshold policies and extended affine policies (Chapter 5).

Chapter 4: Piecewise affine policies. In this chapter, we present a new framework for constructing piecewise affine policies (PAP). In a PAP, we consider pieces $\mathcal{U}_i, i \in [k]$ of $\mathcal{U}$ such that $\mathcal{U}_i \subseteq \mathcal{U}$ and $\mathcal{U}$ is covered by the union of all pieces. For each $\mathcal{U}_i$, we have an affine solution $\boldsymbol{y}(\boldsymbol{h})$ where $\boldsymbol{h} \in \mathcal{U}_i$. PAP are significantly more general than static and affine policies. For problem (1.1), with $\mathcal{U}$ being a polytope, a PAP is known to be optimal. However, the number of pieces can be exponentially large. Moreover, finding the optimal pieces is, in general, an intractable task. In fact, Bertsimas and Caramanis [39] prove that it is NP-hard to construct the optimal pieces, even for pieceiwse policies with two pieces, for two-stage robust linear programs. Here, we focus on the case where he first stage decision $\boldsymbol{x} \in \mathcal{X} = \mathbb{R}^n_+$. We present a tractable framework to construct piecewise affine policies (PAP) with for (1.1) with approximation bounds that improves significantly over affine policies in many settings. Our main contributions in this chapter are as follows.

New Framework for Piecewise affine policy. We present a new framework to efficiently construct a "good" piecewise affine policy for the adjustable robust problem (1.1). As we mentioned earlier, one of the significant challenges in designing a piecewise affine policy arises from the need to construct "good pieces" of the uncertainty set. We suggest a new approach where instead of directly finding an explicit partition of $\mathcal{U}$, we approximate $\mathcal{U}$ with a "simple" set $\hat{\mathcal{U}}$ satisfying the following two properties:

1. the adjustable robust problem (1.1) over $\hat{\mathcal{U}}$ can be solved efficiently,
2. $\hat{\mathcal{U}}$ "dominates" $\mathcal{U}$, i.e., for any $\boldsymbol{h} \in \mathcal{U}$, there exists $\hat{\boldsymbol{h}} \in \hat{\mathcal{U}}$ such that $\boldsymbol{h} \leq \hat{\boldsymbol{h}}$.

Using the uncertainty set $\hat{\mathcal{U}}$ instead of $\mathcal{U}$, the domination property of $\hat{\mathcal{U}}$ preserves the feasibility of the adjustable robust problem. Specifically, we choose $\hat{\mathcal{U}}$ to be a simplex dominating $\mathcal{U}$.

Therefore, the adjustable robust problem (1.1) over $\hat{\mathcal{U}}$ can be solved efficiently since $\hat{\mathcal{U}}$ only has $m+1$ extreme points. We construct a piecewise affine mapping between the uncertainty set $\mathcal{U}$ and the dominating set $\hat{\mathcal{U}}$, i.e. we use a piecewise affine function to map each point $\boldsymbol{h} \in \mathcal{U}$ to a point $\hat{\boldsymbol{h}}$ that dominates $\boldsymbol{h}$. This mapping leads to our piecewise affine policy which is constructed from an optimal adjustable solution over $\hat{\mathcal{U}}$. We show that the performance of our policy is significantly better than the affine policy for many important uncertainty sets both theoretically and numerically.

We elaborate on the two ingredients of designing our piecewise affine policy below, namely, constructing $\hat{\mathcal{U}}$ and the corresponding piecewise map below.

a) **Constructing a dominating uncertainty set**. Our framework is based on choosing an appropriate *dominating simplex* $\hat{\mathcal{U}}$ based on the geometric structure of $\mathcal{U}$. Specifically, $\hat{\mathcal{U}}$ is taken to be a simplex of the following form

$$\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\left(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{v}\right),$$

where $\beta > 0$ and $\boldsymbol{v} \in \mathcal{U}$ are chosen appropriately so that $\hat{\mathcal{U}}$ dominates $\mathcal{U}$. Solving the adjustable robust problem over $\hat{\mathcal{U}}$ gives a feasible solution for problem (1.1) due to the domination property. Moreover, the optimal adjustable solution over $\hat{\mathcal{U}}$ gives a β-approximation for problem (1.1), since $\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\left(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{v}\right) \subseteq \beta \cdot \mathcal{U}$. The approximation bound β is related to a geometric *scaling factor* that represents the Banach-Mazur distance between $\mathcal{U}$ and $\hat{\mathcal{U}}$. We note that $\hat{\mathcal{U}}$ does not necessarily contain $\mathcal{U}$.

b) **The piecewise affine mapping**. We employ the following piecewise affine mapping $\hat{\boldsymbol{h}}(\boldsymbol{h}) = \beta\boldsymbol{v} + (\boldsymbol{h} - \beta\boldsymbol{v})^+$ that maps any $\boldsymbol{h} \in \mathcal{U}$ to a dominating point $\hat{\boldsymbol{h}}$ such that $\boldsymbol{h} \leq \hat{\boldsymbol{h}}$. For any $\boldsymbol{h} \in \mathcal{U}$, $\hat{\boldsymbol{h}}(\boldsymbol{h})$ is contained in the down-monotone completion of $2 \cdot \hat{\mathcal{U}}$. The piecewise affine policy is based on the above piecewise affine mapping and gives a 2β-approximation for problem (1.1). In this policy, $\beta\boldsymbol{v}$ is covered by the static component and $(\boldsymbol{h} - \beta\boldsymbol{v})^+$ is covered by the piecewise linear component of our policy. This is quite analogous to *threshold policies* that are widely used in dynamic optimization. Note that $\hat{\boldsymbol{h}}$ does not necessarily belong to $\hat{\mathcal{U}}$ but is contained in the down-monotone completion of $2 \cdot \hat{\mathcal{U}}$ and therefore, we get an approximation factor of 2β instead of β. We can construct a set-dependent piecewise affine map between $\mathcal{U}$ and $\hat{\mathcal{U}}$ that allows us to construct a piecewise affine policy with a performance bound of β. This bound β is not affected by the scaling introduced in Assumption 1.

Given the dominating set, $\hat{\mathcal{U}}$, our piecewise affine policy can be computed efficiently; in fact, it can be computed even more efficiently than an affine solution over $\mathcal{U}$ in many cases because the adjustable problem over $\hat{\mathcal{U}}$ is a simple LP with only $m + 1$ constraints while the affine problem over $\mathcal{U}$ is a general convex program for general convex uncertainty sets.

Results for Scaled Permutation Invariant (SPI) Sets. The uncertainty set $\mathcal{U}$ is SPI if $\mathcal{U} = \text{diag}(\boldsymbol{\lambda}) \cdot \mathcal{V}$ where $\boldsymbol{\lambda} \in \mathbb{R}_+^m$ and $\mathcal{V}$ is an *invariant set*, i.e., if $\boldsymbol{v} \in \mathcal{V}$, then any permutation of the components of $\boldsymbol{v}$ are also in $\mathcal{V}$. SPI sets include ellipsoids, weighted norm-balls, intersection of norm-balls with budget uncertainty sets and more. SPI sets are commonly used in robust optimization literature and in practice.

We show that for SPI uncertainty set $\mathcal{U}$, it is possible to construct the dominating set $\hat{\mathcal{U}}$ and compute the scaling factor β. In particular, we give an efficiently computable closed-form expression for β and $\boldsymbol{v} \in \mathcal{U}$ that are needed to construct $\hat{\mathcal{U}}$. Consequently, we can efficiently construct our piecewise affine decision rule, having a performance bound 2β.

Using this framework, we provide approximation bounds for the piecewise affine policy that are significantly better than those of the optimal affine policy in [40] for many SPI uncertainty sets. For instance, we show that our policy gives a $O(m^{1/4})$-approximation for the two-stage adjustable robust problem (1.1) with hypersphere uncertainty set as in (4.1), compared to the affine policy in [40] that has an approximation bound of $O(\sqrt{m})$. More generally, the performance bound for our policy for the p-norm ball is $O(m^{\frac{p-1}{p^2}})$ as opposed to $O(m^{\frac{1}{p}})$ given by the affine policy in [40] [1]. Table 1.2 summarizes the above comparisons. We also present computational experiments and observe that our policy also outperforms affine policy in computation time on several examples

[1]**Remark.** We note that in [40], in Tables 1 and 2, there is a typo in the performance bound for affine policies for

of uncertainty sets considered in our experiments including hypersphere, norm-balls and certain polyhedral uncertainty sets. However, we would like to note that our piecewise affine policy does not a generalize affine policy and there are instances where affine policy performs better than our policy. For instance, we observe in our computational experiments that the performance of affine policy is better than our policy for budget of uncertainty sets.

Results for general uncertainty sets. While the dominating set $\hat{\mathcal{U}}$ is given in an efficiently computable closed-form expression for SPI sets, the construction of $\hat{\mathcal{U}}$ for general uncertainty sets requires solving a sequence of MIPs which is computationally much harder than for the case of SPI sets. In Section 4, we give an algorithm for constructing the dominating set $\hat{\mathcal{U}}$, and a piecewise affine policy for general uncertainty set $\mathcal{U}$. Our framework is not necessarily computationally more appealing than computing optimal affine policies. However, we would like to note that in practice these MIPs can be solved efficiently. Moreover, the construction of the dominating set $\hat{\mathcal{U}}$ is independent of the parameters of the adjustable problem and depends only on the uncertainty set, $\mathcal{U}$. Therefore, $\hat{\mathcal{U}}$ can be computed offline and then used to construct the piecewise affine policy efficiently.

We show that our policy gives a $O(\sqrt{m})$-approximation for general uncertainty sets which is same as the worst-case performance bound for affine policy. We also show that the bound of $O(\sqrt{m})$ is tight. In particular, for the budget uncertainty set

$$\mathcal{U} = \left\{ \boldsymbol{h} \in \mathbb{R}^m_+ \;\middle|\; \sum_{i=1}^{m} h_i = \sqrt{m},\ 0 \leq h_i \leq 1 \ \ \forall i \in [m] \right\},$$

the performance bound of our piecewise affine policy is $\Theta(\sqrt{m})$. Furthermore, the bound of $\Theta(\sqrt{m})$ holds even if we consider dominating sets with a polynomial number of extreme points

p-norm balls. According to Theorem 3 in [40], the bound should be

$$\frac{m^{\frac{p-1}{p}} + m}{m^{\frac{p-1}{p}} + m^{\frac{1}{p}}} = O\left(m^{\frac{1}{p}}\right),$$

instead of $\frac{m^{\frac{p-1}{p}}+m}{m^{\frac{1}{p}}+m}$ as mentioned in Table 2 in [40]).

that are significantly more general than a simplex. In Chapter 3, we have shown that affine policies give $O(\frac{\log n}{\log\log n})$-approximation for budget of uncertainty. Therefore, affine policy performs better than our policy for budget of uncertainty sets. While this example shows that the worst-case performance of our policy could be bad, we would like to emphasize that our policy still gives a significantly better approximation than affine policies for many important uncertainty sets including conic sets, and does so in a fraction of computing time (see Section 6.2).

No.	Uncertainty set	Bounds in [40]	Our Bounds
1	$\{\boldsymbol{h} \in \mathbb{R}^m_+ \mid \|\boldsymbol{h}\|_2 \leq 1\}$	$O\left(\sqrt{m}\right)$	$O\left(m^{\frac{1}{4}}\right)$
2	$\left\{\boldsymbol{h} \geq \boldsymbol{0} \mid \sum_{i=1}^m r_i h_i^2 \leq 1\right\}$	$O\left(\sqrt{m}\right)$	$O\left(m^{\frac{1}{4}}\right)$
3	$\left\{\boldsymbol{h} \in \mathbb{R}^m_+ \mid \boldsymbol{h}^T \boldsymbol{\Sigma} \boldsymbol{h} \leq 1\right\}$	—	$O\left(m^{\frac{2}{5}}\right)$
4	$\left\{\boldsymbol{h} \in \mathbb{R}^m_+ \mid \|\boldsymbol{h}\|_p \leq 1\right\}$	$O\left(m^{\frac{1}{p}}\right)$	$O\left(m^{\frac{p-1}{p^2}}\right)$
5	$\left\{\boldsymbol{h} \in \mathbb{R}^m_+ \mid \|\boldsymbol{h}\|_p \leq 1,\ \|\boldsymbol{h}\|_q \leq r\right\}$	$O\left(r^{-1} m^{\frac{1}{q}}\right)$	$O\left(\min\left(r^{\frac{1-p}{p}} m^{\frac{p-1}{pq}}, r^{\frac{1}{q}} m^{\frac{q-1}{q^2}}\right)\right)$
6	$\left\{\boldsymbol{h} \in [0,1]^m \mid \sum_{i=1}^m h_i \leq k\right\}$	$O\left(\frac{k^2+mk}{k^2+m}\right)$	$O\left(\min\left(k, \frac{m}{k}\right)\right)$

Table 1.2: Comparison with performance bounds for affine policies in Bertsimas and Bidkhori [40]. The ellipsoid in Example 3 is assumed to be a permutation invariant set. There is no specialized bound for this Ellipsoid in [40]. For intersection of norm-balls (Example 4 in the table), we assume $m^{\frac{1}{q}-\frac{1}{p}} \geq r \geq 1$. Note that bounds in [40] are not necesserilly tight. For the budget of uncertainty in Example 6, we have shown in Chapter 3 that affine policies gives $O(\frac{\log n}{\log\log n})$ which siginificanly better than the bound in [40].

Chapter 5: Extended affine and threshold policies. In this chapter, we explore new approaches for designing near optimal tractable policies for the two-stage adjustable problem (1.1). In particular, we introduce *extended affine* policies and *threshold polices*. We show that that they significantly improve over the previous known results for approximating (1.1) under some class of uncertainty sets.

Extended Affine policies. An extended affine policy is an affine policy in a lifted space, i.e., instead of restricting the second stage decision to be an affine function of the uncertain parameter $\boldsymbol{h} \in \mathcal{U}$, we first *decompose* $\mathcal{U}$ into several sets and run an affine policy over the new sets. More specifically, we present a framework where we *decompose* an uncertainty set $\mathcal{U}$ into a Minkowski sum of budget of uncertainty sets $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_L$ and define our extended affine policy as the sum of affine policies over $\mathcal{U}_j$ for $j = 1, \ldots, L$. The choice of a decomposition into budget of uncertainty sets is motivated by our results in Chapter 3 on the performance of affine policies for budget of uncertainty sets. In fact, in Chapter 3 , we show that affine policy gives $O(\frac{\log n}{\log\log})$-approximation to (1.1) under budget of uncertainty sets which matches the hardness of approximation for (1.1) and therefore affine policy gives an optimal approximation to (1.1).

More formally, let $\mathcal{U}$ be an uncertainty set. The framework consists of decomposing $\mathcal{U}$ into a a Minkowski sum of small number of budget of uncertainty sets $\mathcal{U}_j$ such that each $\mathcal{U}_j$ is included in $\mathcal{U}$ and $\mathcal{U}$ is within a constant factor from $\mathcal{U}_1 \oplus \mathcal{U}_2 \ldots \oplus \mathcal{U}_L$, i.e.,

- For all $j \in [L]$, $\mathcal{U}_j$ is a budget of uncertainty set.
- For all $j \in [L]$, $\mathcal{U}_j \subseteq \mathcal{U}$.
- $\mathcal{U} \subseteq \gamma \cdot \mathcal{U}_1 \oplus \mathcal{U}_2 \ldots \oplus \mathcal{U}_L$ for some constant γ.

Our extended affine policy is defined as the sum of affine policies over the budgeted sets $\mathcal{U}_j$. We show that this extended affine policy gives $O\left(\frac{\gamma L \log n}{\log\log n}\right)$-approximation to (1.1), i.e.,

$$\frac{1}{\gamma} z_{\mathsf{AR}}(\mathcal{U}) \leq \sum_{j=1}^{L} z_{\mathsf{Aff}}(\mathcal{U}_j) \leq O\left(\frac{L \log n}{\log\log n}\right) \cdot z_{\mathsf{AR}}(\mathcal{U}),$$

where $z_{\mathsf{Aff}}(\mathcal{U}_j)$ is the cost of the optimal affine policy over $\mathcal{U}_j$.

We give an explicit construction of this decomposition for important class of uncertainty sets that can be computed efficiently. We show that our extended affine policy gives $O(\frac{\log n \log m}{\log\log n})$-approximation for the important class of permutation invariant sets that includes hypersphere and q-norm balls. This approximation bound improves significantly over the previous results in the literature, for instance the best known bound in the litterature is $O\left(m^{\frac{1}{4}}\right)$ for hypersphere and $O\left(m^{\frac{q-1}{q^2}}\right)$ [41] for q-norm balls. To the best of our knowledge, the approximation bounds in this chapter are the first logarithmic approximation bounds for (1.1) under conic uncertainty sets.

Threshold policies. In the second part of this chapter, our goal is to characterize the structure of near-optimal solutions for (1.1). In particular, we present threshold policies. These are particular class of piecewise affine policies where the second-stage decision is restricted to be of the form:

$$\boldsymbol{y}(\boldsymbol{h}) = \sum_{i=1}^{m} (h_i - \theta_i)^+ \boldsymbol{v}_i + \boldsymbol{q}.$$

Here, $\boldsymbol{\theta} \in \mathbb{R}^m_+$ is the threshold parameter, $\boldsymbol{q} \in \mathbb{R}^n_+$ and for all $i \in [m]$ $\boldsymbol{v}_i \in \mathbb{R}^n_+$. Threshold policies are widely used in practice in many settings and applications (see for instance [42]). They are highly interpretable and easy to implement in practice. However computing optimal threshold policies is often a hard problem. Here, our goal is not to compute optimal threshold policies, but to analyze the structure of a near-optimal policy for (1.1) and show that it could be captured by a threshold policy. In particular, based on insights from the construction of our extended affine policy, we show by construction the existence of threshold policies that gives $O(\log n + \log m)$ approximation for (1.1) for hypersphere and q-norm ball uncertainty sets and give $O(\log n \log m)$-approximation for the general class of permutation invariant sets. The construction can be computed efficiently, however, it needs to guess the value or an approximate value of OPT. These bounds almost match the hardness of (1.1) and therefore the structure a near-optimal solution for (1.1), could be given by a threshold policy.

1.2 Two-stage robust optimization with packing constraints

In Chapter 6, we consider two-stage adjustable robust linear optimization problem under uncertain packing constraints. The problem is given by

$$\begin{aligned} z_{\mathsf{AR}}(\mathcal{U}) = \max \;\; & \boldsymbol{c}^T \boldsymbol{x} + \min_{\boldsymbol{B} \in \mathcal{U}} \max_{\boldsymbol{y}(\boldsymbol{B})} \boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{B}) \\ & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{B}) \leq \boldsymbol{h} \\ & \boldsymbol{x} \in \mathbb{R}^{n_1}_+, \; \boldsymbol{y}(\boldsymbol{B}) \in \mathbb{R}^{n_2}_+, \end{aligned} \tag{1.7}$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n_1}$, $\boldsymbol{c} \in \mathbb{R}_+^{n_1}$, $\boldsymbol{d} \in \mathbb{R}_+^{n_2}$, $\boldsymbol{h} \in \mathbb{R}_+^m$. Also, $\mathcal{U} \subseteq R_+^{m \times n_2}$ is a full dimensional compact convex *down-monotone* uncertainty set in the non-negative orthant. Following Bertsimas et al. [43], we can assume without loss of generality that $\mathcal{U}$ is *down-monotone* and $n_1 = n_2 = n$ (A set $\mathcal{S} \subseteq \mathbb{R}_+^n$ is *down-monotone* if $\boldsymbol{s} \in \mathcal{S}$, $\boldsymbol{t} \in \mathbb{R}_+^n$ and $\boldsymbol{t} \leq \boldsymbol{s}$ implies $\boldsymbol{t} \in \mathcal{S}$). Note that $\boldsymbol{x}$ represents the first-stage decisions and $\boldsymbol{y}(\boldsymbol{B})$ represents the second-stage decisions after observing the uncertain constraint matrix $\boldsymbol{B} \in \mathcal{U}$.

The above formulation models many interesting applications including revenue management and resource allocation problems with uncertain demand. For instance, in a resource allocation application, the right hand side $\boldsymbol{h}$ can model the fixed resource capacities and the uncertain coefficients in $\boldsymbol{B}$ model the uncertain requirements of resources for demand. The goal is to find an optimal allocation of resources that maximizes the worst case profit (see Wiesemann [44]). When $m = 1$, the above problem reduces to a fractional knapsack problem with uncertain item sizes. The stochastic version of the knapsack problem has been widely studied in the literature (see Dean et al. [45], Goel and Indyk [46], Goyal and Ravi [47]).

In general, it is intractable to compute an adjustable robust solution for (1.7). In fact, Awasthi et al. [48] show that the two-stage adjustable robust problem (1.7) is $\Omega(\log n)$-hard to approximate if the uncertainty set of constraint coefficients belongs to the non-negative orthant. In other words, there is no polynomial time algorithm that approximates the optimal adjustable solution within a factor better than $\log n$. Therefore, the goal is to construct approximate policies with good performance. A static solution approach, where we give a single solution feasible for all scenarios, has been widely studied in the literature. We can formulate the static robust optimization problem $\Pi_{\mathsf{Rob}}(\mathcal{U})$ to approximate (1.7) as follows.

$$
\begin{aligned}
z_{\mathsf{Rob}}(\mathcal{U}) = \max \ & \boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{d}^T \boldsymbol{y} \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \leq \boldsymbol{h} \ \ \forall \boldsymbol{B} \in \mathcal{U} \\
& \boldsymbol{x} \in \mathbb{R}_+^n, \ \boldsymbol{y} \in \mathbb{R}_+^n.
\end{aligned}
\tag{1.8}
$$

An optimal static solution can be computed efficiently for large class of problems (see Bertsimas et al. [15], Ben-Tal et al. [16]). Ben-Tal and Nemirovski [9] show that a static solution is optimal for (1.7) if the uncertainty set is constraint-wise where each constraint is selected independently from a compact convex set $\mathcal{U}_i$ (i.e. $\mathcal{U}$ is a Cartesian product of $\mathcal{U}_i$, $i = 1, \ldots, m$). Bertsimas et al.[43] generalize the result of [9] and show that a static solution is near optimal for several interesting families of $\mathcal{U}$. In particular, they give a tight characterization on the performance of the static solution related to the measure of non-convexity of a transformation of the uncertainty set $\mathcal{U}$. While a static solution provides a good approximation in many cases, it can be as bad as a factor m away from the optimal adjustable solution in general.

Piecewise static policies is another solution approach that has been studied in the literature. A piecewise static policy is a generalization of the static policy where the uncertainty set is divided into several pieces and we specify a static policy for each piece. Bertsimas and Caramanis [39] consider a piecewise static solution approach (also referred to as *finite K-adaptability*) where they propose a hierarchy of increasing adaptability that bridges the gap between the static robust formulation, and the fully adaptable formulation. Hanasusanto et al. [31] consider a K-adaptable solution approach for two-stage robust integers optimization problems.

1.2.1 Summary of contributions of Chapter 6

In Chapter 6, we consider the piecewise static solution approach for two-stage adjustable problem with capacity constraints (1.7). In particular, we consider a piecewise policy with p pieces (or subsets): $\mathcal{U}_1, \ldots, \mathcal{U}_p$ of $\mathcal{U}$ such that

$$\mathcal{U} = \bigcup_{1 \le i \le p} \mathcal{U}_i,$$

where each $\mathcal{U}_i$ is convex, compact and down-monotone uncertainty subset. Note that $\mathcal{U}_i$ are not necessarily disjoint. We can formulate the two-stage piecewise robust linear optimization problem as follows:

$$z_{\mathsf{PR}}(\mathcal{U}_1,\ldots,\mathcal{U}_p) = \max \;\; \boldsymbol{c}^T\boldsymbol{x} + \; \min(\boldsymbol{d}^T\boldsymbol{y}_1, \boldsymbol{d}^T\boldsymbol{y}_2, \ldots, \boldsymbol{d}^T\boldsymbol{y}_p)$$
$$\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}_i\boldsymbol{y}_i \leq \boldsymbol{h} \qquad \forall i \in [p]\ ,\ \forall \boldsymbol{B}_i \in \mathcal{U}_i \tag{1.9}$$
$$\boldsymbol{x} \in \mathbb{R}^n_+,\ \boldsymbol{y}_i \in \mathbb{R}^n_+ \ \ \forall i \in [p].$$

We show that the performance of the optimal piecewise static policy for given pieces is related to the maximum of the measures of non-convexity of transformations of the pieces $\mathcal{U}_i$; thereby extending the bound in [43] for piecewise static policies. Note that if the pieces $\mathcal{U}_i$ are given explicitly, we can efficiently compute an optimal piecewise static policy provide we can solve linear optimization over each $\mathcal{U}_i$ efficiently. However, one of the main challenges in designing a good piecewise static policy, is to construct good pieces of the uncertainty set. In fact, Bertsimas and Caramanis [39] show that it is NP-hard to construct the optimal pieces for piecewise policies with only two pieces for two-stage robust linear programs in general.

Our main contribution in this chapter is to show that even if we ignore the computational complexity of computing optimal pieces, surprisingly the performance of piecewise static policies with a polynomial number of pieces is not significantly better than a static policy in general. In particular, we show that there is no piecewise static policy with polynomial number of pieces that gives an approximation bound better than $O(m^{1-\epsilon})$ for any $\epsilon > 0$ for general uncertainty sets $\mathcal{U} \subseteq R^{m\times n}_+$ where the approximation bound for the static policy is m. We prove this by constructing a family of instances of $\mathcal{U}$ for any $\epsilon > 0$, such that the performance of the static policy is m and the performance of any piecewise policy with polynomial number of pieces is $\Omega(m^{1-\epsilon})$. Our proof is based on a combinatorial argument and structural results about piecewise static policies.

Chapter 2: Beyond worst-case: a probabilistic analysis of affine policies

2.1 Introduction

Affine policies (or control) are widely used as a solution approach in dynamic optimization where computing an optimal adjustable solution is usually intractable. While the worst case performance of affine policies can be significantly bad, the empirical performance is observed to be near-optimal for a large class of problem instances. For instance, in the two-stage dynamic robust optimization problem with linear covering constraints and uncertain right hand side (1.1) , the worst-case approximation bound for affine policies is $O(\sqrt{m})$ that is also tight (see Bertsimas and Goyal [18]), whereas observed empirical performance is near-optimal. In this chapter, we aim to address this stark-contrast between the worst-case and the empirical performance of affine policies. In particular, we show that with high probability affine policies give a good approximation for two-stage dynamic robust optimization problems on random instances generated from a large class of distributions; thereby, providing a theoretical justification of the observed empirical performance. The approximation bound depends on the distribution, but it is significantly better than the worst-case bound for a large class of distributions.

The rest of this chapter is organized as follows. In Section 2.2, we present our results on the performance of affine policies for random instances and show that affine policies give with high probability *good* approximation to (1.1) for a large class of distributions. In Section 2.3, we present a class of distributions and bad instances where affine policies perform poorly and match the worst-case deterministic bound. Finally, we present a computational study to test affine policies on random instances in Section 2.4.

2.2 Random instances of two-stage robust optimization problems

Recall the two-stage adjustable problem (1.1) ,

$$
\begin{aligned}
z_{\mathsf{AR}}(\boldsymbol{c},\boldsymbol{d},\boldsymbol{A},\boldsymbol{B},\mathcal{U}) = \min_{\boldsymbol{x}} \ & \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \min_{\boldsymbol{y}(\boldsymbol{h})} \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{h}) \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) \ \geq \ \boldsymbol{h}, \quad \forall \boldsymbol{h} \in \mathcal{U} \\
& \boldsymbol{x} \in \mathcal{X} \\
& \boldsymbol{y}(\boldsymbol{h}) \in \mathbb{R}^n_+, \quad \forall \boldsymbol{h} \in \mathcal{U}.
\end{aligned}
$$

In this section, we theoretically characterize the performance of affine policies for random instances of (1.1). In particular, we consider the two-stage problem where coefficients of constraint matrix $\boldsymbol{B}$ are random and analyze the performance of affine policies for a large class of distributions. Our analysis of the performance of affine policies does not depend on the structure of first stage constraint matrix $\boldsymbol{A}$, cost $\boldsymbol{c}$ or the choice of uncertainty set $\mathcal{U}$. We assume without loss of generality that $\boldsymbol{c} = \boldsymbol{e}$ and $\boldsymbol{d} = \bar{d} \cdot \boldsymbol{e}$ (by appropriately scaling $\boldsymbol{A}$ and $\boldsymbol{B}$). Here, $\bar{d}$ can interpreted as the inflation factor for costs in the second-stage. Finally, we assume in this chapter that the first stage decision $\boldsymbol{x}$ belongs to a polyhedral cone $\mathcal{X}$, i.e., if $\boldsymbol{x} \in \mathcal{X}$ then $\alpha\boldsymbol{x} \in \mathcal{X}$ for any $\alpha > 0$ (for example $\mathcal{X} = \mathbb{R}^n_+$).

Therefore, we restrict our attention only to the distribution of coefficients of the second stage matrix $\boldsymbol{B}$. We will use the notation $\tilde{\boldsymbol{B}}$ to emphasize that $\boldsymbol{B}$ is random. For simplicity, we refer to $z_{\mathsf{AR}}(\boldsymbol{c},\boldsymbol{d},\boldsymbol{A},\boldsymbol{B},\mathcal{U})$ as $z_{\mathsf{AR}}(\boldsymbol{B})$ and to $z_{\mathsf{Aff}}(\boldsymbol{c},\boldsymbol{d},\boldsymbol{A},\boldsymbol{B},\mathcal{U})$ as $z_{\mathsf{Aff}}(\boldsymbol{B})$.

We first consider the case when the columns of $\tilde{\boldsymbol{B}}$, namely $\tilde{\boldsymbol{B}}_j$ for $j \in [n]$, is distributed according to a multivariate distribution $\mathcal{F}_j$ with bounded support in $[0,b]^m$ (for some constant b), independent of other columns. We compare the performance of affine policies with respect to the optimal dynamic solution and present an approximation bound that depends only on the distribution of $\tilde{\boldsymbol{B}}$ and holds for any uncertainty set $\mathcal{U}$. In particular, we have the following theorem.

Theorem 2.2.1 (Distributions with bounded support)**.** *Consider the two-stage adjustable problem*

(1.1) *where* $\tilde{\boldsymbol{B}}_j$ *for* $j \in [n]$ *is distributed according to a multivariate distribution,* $\mathcal{F}_j$ *with bounded support in* $[0,b]^m$ *(for some constant b), independent of other columns. Let* $\mathbb{E}[\tilde{B}_{ij}] = \mu_{ij}\ \forall i \in [m]\ \forall j \in [n]$*. For n and m sufficiently large, we have with probability at least* $1 - \frac{1}{m}$*,*

$$z_{\mathsf{AR}}(\tilde{\boldsymbol{B}}) \le z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}}) \le \frac{b}{\mu(1-\epsilon)} \cdot z_{\mathsf{AR}}(\tilde{\boldsymbol{B}})$$

where $\mu = \min_{i\in[m]} \frac{1}{n}\sum_{j=1}^n \mu_{ij}$ *and* $\epsilon = \frac{b}{\mu}\sqrt{\frac{\log m}{n}}$*.*

For the special case where $\tilde{B}_{ij}$ are i.i.d. according to a bounded distribution with support in $[0,b]$. We have the following corollary.

Corollary 2.2.2. *Consider the two-stage adjustable problem* (1.1) *where* $\tilde{B}_{ij}$ *are i.i.d. according to a bounded distribution with support in* $[0,b]$ *and expectation* μ*. For n and m sufficiently large, we have with probability at least* $1 - \frac{1}{m}$*,*

$$z_{\mathsf{AR}}(\tilde{\boldsymbol{B}}) \le z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}}) \le \frac{b}{\mu(1-\epsilon)} \cdot z_{\mathsf{AR}}(\tilde{\boldsymbol{B}})$$

where $\epsilon = \frac{b}{\mu}\sqrt{\frac{\log m}{n}}$*.*

The above theorem and corollary show that for sufficiently large values of m and n, the performance of affine policies is at most b/μ times the performance of an optimal adjustable solution. This shows that affine policies give a good approximation (and significantly better than the worst-case bound of $O(\sqrt{m})$) for many important distributions. We present some examples below.

Example 1. [Uniform distribution] Suppose for all $i \in [m]$ and $j \in [n]$ $\tilde{B}_{ij}$ are i.i.d. uniform in $[0,1]$. Then $\mu = 1/2$ and from Corollary 2.2.2 we have with probability at least $1 - 1/m$,

$$z_{\mathsf{AR}}(\tilde{\boldsymbol{B}}) \le z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}}) \le \frac{2}{1-\epsilon} \cdot z_{\mathsf{AR}}(\tilde{\boldsymbol{B}})$$

where $\epsilon = 2\sqrt{\log m/n}$. Therefore, for sufficiently large values of n and m affine policy gives a 2-approximation to the adjustable problem in this case. Note that the approximation bound of 2 is a conservative bound and the empirical performance is significantly better. We demonstrate this in our numerical experiments.

Example 2. [Bernoulli distribution] Suppose for all $i \in [m]$ and $j \in [n]$, $\tilde{B}_{ij}$ are i.i.d. according to a Bernoulli distribution of parameter p. Then $\mu = p$, $b = 1$ and from Corollary 2.2.2 we have with probability at least $1 - \frac{1}{m}$,

$$z_{\mathsf{AR}}(\tilde{\boldsymbol{B}}) \leq z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}}) \leq \frac{1}{p(1-\epsilon)} \cdot z_{\mathsf{AR}}(\tilde{\boldsymbol{B}})$$

where $\epsilon = \frac{1}{p}\sqrt{\frac{\log m}{n}}$. Therefore for constant p, affine policy gives a constant approximation to the adjustable problem (for example 2-approximation for $p = 1/2$).

Note that these performance bounds are in stark contrast with the worst case performance bound $O(\sqrt{m})$ for affine policies which is tight. For these random instances, the performance is significantly better. We would like to note that the above distributions are very commonly used to generate instances for testing the performance of affine policies and exhibit good empirical performance. Here, we give a theoretical justification of the good empirical performance of affine policies on such instances, thereby closing the gap between worst case bound of $O(\sqrt{m})$ and observed empirical performance.

While the approximation bound in Theorem 2.2.1 leads to a good approximation for many distributions, the ratio b/μ can be significantly large in general. We can tighten the analysis by using the concentration properties of distributions and can extend the analysis for the case of distributions with sub-gaussian tails. In particular, we consider the case where $\tilde{\boldsymbol{B}}_j$ is generated according to a distribution with sub-gaussian tails and show a logarithmic approximation bound for affine policies. Note that we assume that the parameters of the distribution are independent of the problem dimensions. We have the following theorem.

Theorem 2.2.3 (Distributions with sub-gaussian tails)**.** *Suppose* $\forall j \in [n]$, $\tilde{\boldsymbol{B}}_j = |\tilde{\boldsymbol{G}}_j|$ *such that* $\tilde{\boldsymbol{G}}_j$ *is a sub-Gaussian, independent of* $\tilde{\boldsymbol{G}}_i$*, for all* $i \neq j$*. For* n *and* m *sufficiently large, we have with* *probability at least* $1 - \frac{1}{m}$,

$$z_{\mathsf{AR}}(\tilde{\boldsymbol{B}}) \leq z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}}) \leq \kappa \cdot z_{\mathsf{AR}}(\tilde{\boldsymbol{B}})$$

where $\kappa = O\left(\sqrt{\log m + \log n}\right)$.

We would like to note that the bound of $O\left(\sqrt{\log m + \log n}\right)$ depends on the dimension of the problem unlike the case of uniform bounded distributions. But, it is significantly better than the worst-case of $O(\sqrt{m})$ [18] for general instances. Furthermore, this bound holds for all uncertainty sets. We would like to note though that the bounds are not necessarily tight. In fact, in our numerical experiments where the uncertainty set is a *budget of uncertainty*, we observe that the performance is much better than the bounds. We discuss the intuition and the proofs of Theorem 2.2.1 and Theorem 2.2.3 in the following subsections.

2.2.1 Preliminaries

In order to prove Theorem 2.2.1 and Theorem 2.2.3, we need to introduce certain preliminary results. First, to develop intuition, let us consider the case of polyhedral uncertainty set $\mathcal{U}$, i.e.,

$$\mathcal{U} = \{\boldsymbol{h} \in \mathbb{R}^m_+ \mid \boldsymbol{R}\boldsymbol{h} \leq \boldsymbol{r}\} \tag{2.1}$$

where $\boldsymbol{R} \in \mathbb{R}^{L\times m}$ and $\boldsymbol{r} \in \mathbb{R}^L$. This is a fairly general class of uncertainty sets that includes many commonly used sets such as *box uncertainty* sets and *budget of uncertainty* sets. In section 2.2.4, we sketch the extension of our results to general convex uncertainty sets such as ellipsoids.

We first introduce the following formulation for the adjustable problem (1.1) based on ideas in Bertsimas and de Ruiter [17].

$$\begin{aligned} z_{\mathsf{d-AR}}(\boldsymbol{B}) = \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^T\boldsymbol{x} + &\max_{\boldsymbol{w}\in\mathcal{W}} \min_{\boldsymbol{\lambda}(\boldsymbol{w})} -(\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{w} + \boldsymbol{r}^T\boldsymbol{\lambda}(\boldsymbol{w}) \\ &\boldsymbol{R}^T\boldsymbol{\lambda}(\boldsymbol{w}) \;\geq\; \boldsymbol{w}, \quad \forall \boldsymbol{w} \in \mathcal{W} \\ &\boldsymbol{\lambda}(\boldsymbol{w}) \in \mathbb{R}^L_+, \; \forall \boldsymbol{w} \in \mathcal{W} \end{aligned} \tag{2.2}$$

where the set $\mathcal{W}$ is defined as

$$\mathcal{W} = \{\boldsymbol{w} \in \mathbb{R}^m_+ \mid \boldsymbol{B}^T \boldsymbol{w} \leq \boldsymbol{d}\}. \tag{2.3}$$

We show that the above problem is an equivalent formulation of (1.1).

Lemma 2.2.4. *Let $z_{\mathsf{AR}}(\boldsymbol{B})$ be as defined in* (1.1) *and $z_{\mathsf{d-AR}}(\boldsymbol{B})$ as defined in* (2.2)*. Then,*

$$z_{\mathsf{AR}}(\boldsymbol{B}) = z_{\mathsf{d-AR}}(\boldsymbol{B}).$$

The proof follows from [17]. For completeness, we present it in Appendix B.1. Reformulation (2.2) can be interpreted as a new two-stage adjustable problem over *dualized* uncertainty set $\mathcal{W}$ and decision $\boldsymbol{\lambda}(\boldsymbol{w})$. Following [17], we refer to (2.2) as the *dualized* formulation and to (1.1) as the *primal* formulation. Bertsimas and de Ruiter [17] show that even the affine approximations of (1.1) and (2.2) (where recourse decisions are restricted to be affine functions of respective uncertainties) are equivalent. In particular, we have the following Lemma which is a restatement of Theorem 2 in [17].

Lemma 2.2.5. (Theorem 2 in Bertsimas and de Ruiter [17]) *Let $z_{\mathsf{d-Aff}}(\boldsymbol{B})$ be the objective value when $\boldsymbol{\lambda}(\boldsymbol{w})$ is restricted to be affine function of $\boldsymbol{w}$ and $z_{\mathsf{Aff}}(\boldsymbol{B})$ as defined in* (1.2)*. Then,*

$$z_{\mathsf{d-Aff}}(\boldsymbol{B}) = z_{\mathsf{Aff}}(\boldsymbol{B}).$$

Bertsimas and Goyal [18] show that affine policy is optimal for the adjustable problem (1.1) when the uncertainty set $\mathcal{U}$ is a simplex. In fact, optimality of affine policies for simplex uncertainty sets holds for more general formulation than considered in [18]. In particular, we have the following lemma.

Lemma 2.2.6. *Suppose the set $\mathcal{W}$ is a simplex, i.e. a convex combination of $m+1$ affinley independent points, then affine policy is optimal for the adjustable problem* (2.2)*, i.e.* $z_{\mathsf{d-Aff}}(\boldsymbol{B}) = z_{\mathsf{d-AR}}(\boldsymbol{B})$.

The proof proceeds along similar lines as in [18]. For completeness, we provide it in Appendix B.1. In fact, if the uncertainty set is not simplex but can be approximated by a simplex within a small scaling factor, affine policies can still be shown to be a good approximation, in particular we have the following lemma.

Lemma 2.2.7. *Denote $\mathcal{W}$ the dualized uncertainty set as defined in* (2.3) *and suppose there exists a simplex $\mathcal{S}$ and $\kappa \geq 1$ such that $\mathcal{S} \subseteq \mathcal{W} \subseteq \kappa \cdot \mathcal{S}$. Therefore,*

$$z_{\mathsf{d-AR}}(\boldsymbol{B}) \leq z_{\mathsf{d-Aff}}(\boldsymbol{B}) \leq \kappa \cdot z_{\mathsf{d-AR}}(\boldsymbol{B}).$$

Furthermore,

$$z_{\mathsf{AR}}(\boldsymbol{B}) \leq z_{\mathsf{Aff}}(\boldsymbol{B}) \leq \kappa \cdot z_{\mathsf{AR}}(\boldsymbol{B}).$$

The proof of Lemma 2.2.7 is presented in Appendix B.1.

2.2.2 Proof of Theorem 2.2.1

We consider instances of problem (1.1) where the columns $\tilde{\boldsymbol{B}}_j$ are independently generated according to bounded distributions with support in $[0, b]^m$. Let $\mathbb{E}[\tilde{B}_{ij}] = \mu_{ij}$ for all $i \in [m], j \in [n]$ and

$$\mu = \min_{i \in [m]} \frac{1}{n} \sum_{j=1}^{n} \mu_{ij}.$$

Denote the dualized uncertainty set

$$\tilde{\mathcal{W}} = \left\{ \boldsymbol{w} \in \mathbb{R}_+^m \mid \tilde{\boldsymbol{B}}^T \boldsymbol{w} \leq \bar{d} \cdot \boldsymbol{e} \right\}.$$

Our performance bound is based on showing that $\tilde{\mathcal{W}}$ can be sandwiched between two simplicies with a small scaling factor. In particular, consider the following simplex,

$$\mathcal{S} = \left\{ \boldsymbol{w} \in \mathbb{R}_+^m \;\middle|\; \sum_{i=1}^{m} w_i \leq \frac{\bar{d}}{b} \right\}. \tag{2.4}$$

We will show that

$$\mathcal{S} \subseteq \tilde{\mathcal{W}} \subseteq \frac{b}{\mu(1-\epsilon)} \cdot \mathcal{S}$$

with probability at least $1 - \frac{1}{m}$ where $\epsilon = \frac{b}{\mu}\sqrt{\frac{\log m}{n}}$. First, we show that $\mathcal{S} \subseteq \tilde{\mathcal{W}}$. Consider any $\boldsymbol{w} \in \mathcal{S}$. For $j = 1, \ldots, n$, we have

$$\sum_{i=1}^{m} \tilde{B}_{ij} w_i \leq b \sum_{i=1}^{m} w_i \leq \bar{d}.$$

The first inequality holds because all components of $\tilde{\boldsymbol{B}}$ are upper bounded by b and the second one follows from $\boldsymbol{w} \in \mathcal{S}$. Hence, we have $\tilde{\boldsymbol{B}}^T \boldsymbol{w} \leq \bar{d}\boldsymbol{e}$ and consequently $\mathcal{S} \subseteq \tilde{\mathcal{W}}$.

Now, we show that the other inclusion holds with high probability. Consider any $\boldsymbol{w} \in \tilde{\mathcal{W}}$. We have $\tilde{\boldsymbol{B}}^T \boldsymbol{w} \leq \bar{d} \cdot \boldsymbol{e}$. Summing up all the inequalities and dividing by n, we get

$$\sum_{i=1}^{m} \left(\frac{\sum_{j=1}^{n} \tilde{B}_{ij}}{n} \right) \cdot w_i \leq \bar{d}. \tag{2.5}$$

The columnds of $\boldsymbol{B}$ are independent, hence using Hoeffding's inequality [49] with $\tau = b\sqrt{\frac{\log m}{n}}$ (see Appendix B.2), we have for all $i \in [m]$,

$$\mathbb{P}\left(\frac{\sum_{j=1}^{n} \tilde{B}_{ij}}{n} - \mu_i \geq -\tau \right) \geq 1 - \exp\left(\frac{-2n\tau^2}{b^2} \right) = 1 - \frac{1}{m^2}$$

where $\mu_i = \frac{1}{n}\sum_{j=1}^{n} \mu_{ij}$. Then, a union bound over $i = 1, \ldots, m$ gives us

$$\mathbb{P}\left(\frac{\sum_{j=1}^{n} \tilde{B}_{ij}}{n} \geq \mu_i - \tau \;\; \forall i \in [m] \right) \geq 1 - \sum_{i=1}^{m} \mathbb{P}\left(\frac{\sum_{j=1}^{n} \tilde{B}_{ij}}{n} < \mu_i - \tau \right) \geq 1 - \sum_{i=1}^{m} \frac{1}{m^2} = 1 - \frac{1}{m}.$$

Therefore, with probability at least $1 - \frac{1}{m}$, we have

$$\sum_{i=1}^{m} w_i \leq \sum_{i=1}^{m} \frac{1}{\mu_i - \tau} \left(\frac{\sum_{j=1}^{n} \tilde{B}_{ij}}{n} \right) \cdot w_i \leq \frac{1}{\min_{i \in [m]} \mu_i - \tau} \cdot \sum_{i=1}^{m} \left(\frac{\sum_{j=1}^{n} \tilde{B}_{ij}}{n} \right) \cdot w_i \leq \frac{\bar{d}}{\mu - \tau} = \frac{b}{\mu(1-\epsilon)} \cdot \frac{\bar{d}}{b}$$

where the last inequality follows from (2.5). Note that for m sufficiently large , we have $\mu - \tau > 0$. Then, $\boldsymbol{w} \in \frac{b}{\mu(1-\epsilon)} \cdot \mathcal{S}$ for any $\boldsymbol{w} \in \tilde{\mathcal{W}}$. Consequently with probability at least $1 - 1/m$, we have

$$\mathcal{S} \subseteq \tilde{\mathcal{W}} \subseteq \frac{b}{\mu(1-\epsilon)} \cdot \mathcal{S}.$$

Finally, we apply the result of Lemma 2.2.7 to conclude. □

2.2.3 Proof of Theorem 2.2.3

Consider instances of problem (1.1) where the columns $\tilde{\boldsymbol{B}}_j$ are independently generated according to distributions with sub-gaussian tails. In particular, we have for all i, j, $\tilde{B}_{ij} = |\tilde{G}_{ij}|$ where $\tilde{G}_{ij}$ is a sub-Gaussian random variable. Denote

$$\tilde{\mathcal{W}} = \{\boldsymbol{w} \in \mathbb{R}_+^m \mid \tilde{\boldsymbol{B}}^T \boldsymbol{w} \leq \bar{d} \cdot \boldsymbol{e}\}.$$

Our goal is to sandwich $\tilde{\mathcal{W}}$ between two simplicies and use Lemma 2.2.7. Since $\tilde{G}_{ij}$ has a sub-gaussian tail, there exists positive constants C and v_{ij} such that for any $t > 0$,

$$\mathbb{P}\left(|\tilde{G}_{ij}| \geq t\right) \leq Ce^{-v_{ij}t^2}.$$

Therefore,

$$\begin{aligned}\mathbb{P}\left(\tilde{B}_{ij} \leq \sqrt{\frac{2\log(mn)}{v_{ij}}}\right) &= 1 - \mathbb{P}\left(|\tilde{G}_{ij}| > \sqrt{\frac{2\log(mn)}{v_{ij}}}\right) \\ &\geq 1 - C\exp\left(-2\log(mn)\right) = 1 - \frac{C}{(mn)^2}.\end{aligned}$$

Denote

$$\kappa = \max_{i,j}\left(\sqrt{\frac{2\log(mn)}{v_{ij}}}\right).$$

We have $\kappa = O\left(\sqrt{\log m + \log n}\right)$ because v_{ij} are positive constant independent of the dimensions m and n of the problem. Therefore by taking a union bound over $i \in [m]$ and $j \in [n]$ we get,

$$\mathbb{P}\left(\tilde{B}_{ij} \le \kappa \ \ \forall i \in [m], \forall j \in [n]\right) \ge 1 - \frac{C}{mn}.$$

Consider the following simplex

$$\mathcal{S} = \{\boldsymbol{w} \in \mathbb{R}_+^m \mid \sum_{i=1}^m w_i \le \bar{d}\}.$$

For any $w \in \mathcal{S}$, we have with probability at least $1 - \frac{C}{mn}$,

$$\sum_{i=1}^m \tilde{B}_{ij} w_i \le \kappa \sum_{i=1}^m w_i \le \kappa \cdot \bar{d} \qquad \forall j \in [n].$$

Hence, with probability at least $1 - \frac{C}{mn}$ we have, $\mathcal{S} \subseteq \kappa \cdot \tilde{\mathcal{W}}$. Now, we want to find a simplex that includes $\tilde{\mathcal{W}}$. We follow a similar approach to the proof of Theorem 2.2.1. Consider any $\boldsymbol{w} \in \tilde{\mathcal{W}}$. We have similarly to equation (2.5)

$$\sum_{i=1}^m \left(\frac{\sum_{j=1}^n \tilde{B}_{ij}}{n}\right) \cdot w_i \le \bar{d}. \tag{2.6}$$

We have the following concentration inequality for non-negative random variables,

$$\mathbb{P}\left(\frac{\sum_{j=1}^n \tilde{B}_{ij}}{n} \ge \mu_i - \tau_i\right) \ge 1 - \exp\left(\frac{-n\tau_i^2}{2\sigma_i^2}\right) = 1 - \frac{1}{m^2}$$

where $\tau_i = 2\sigma_i \sqrt{\frac{\log m}{n}}$, $\mu_i = \frac{1}{n}\sum_{j=1}^n \mathbb{E}[\tilde{B}_{ij}]$ and $\sigma_i^2 = \max_j \mathsf{Var}[\tilde{B}_{ij}]$. Then, a union bound over $i \in [m]$ gives

$$\mathbb{P}\left(\frac{\sum_{j=1}^n \tilde{B}_{ij}}{n} \ge \mu_i - \tau_i, \ \ \forall i \in [m]\right) \ge 1 - \frac{1}{m},$$

which implies

$$\mathbb{P}\left(\frac{\sum_{j=1}^{n}\tilde{B}_{ij}}{n} \geq \kappa', \;\; \forall i \in [m]\right) \geq 1 - \frac{1}{m}.$$

where $\kappa' = \max_{i\in[m]}(\mu_i - \tau_i)$. Therefore, combining this result with inequality (2.6), we have with probability at least $1 - \frac{1}{m}$, $\tilde{\mathcal{W}} \subseteq \frac{1}{\kappa'} \cdot \mathcal{S}$. Denote, $\mathcal{S}' = \frac{1}{\kappa}\mathcal{S}$. We have shown that with probability at least $1 - C/mn$, $\mathcal{S}' \subseteq \tilde{\mathcal{W}}$. Therefore, we have with probability at least $1 - \frac{1}{m}$,

$$\mathcal{S}' \subseteq \tilde{\mathcal{W}} \subseteq \frac{\kappa}{\kappa'} \cdot \mathcal{S}'$$

where

$$\frac{\kappa}{\kappa'} = O\left(\sqrt{\log m + \log n}\right),$$

for sufficiently large values of m and n. We finally use Lemma 2.2.7 to conclude.

2.2.4 Extension to general convex uncertainty sets

In this section, we show that our results of Theorem 2.2.1 and Theorem 2.2.3 hold as well for general convex uncertainty sets $\mathcal{U}$ including ellipsoids and norm-ball sets that are widely used in robust optimization. This is based on approximating a convex uncertainty set by a polyhedral set (possibly given by an exponential number of inequalities). In fact, in Section 2.2.2 and Section 2.2.3, we prove Theorem 2.2.1 and Theorem 2.2.3 for the case of polyhedral uncertainty set $\mathcal{U}$. Note that the approximation bounds are independent from the description of $\mathcal{U}$ and depend only on the distribution of $\tilde{\boldsymbol{B}}$.

Now, consider a genral convex uncertainty set $\mathcal{U} \subseteq \mathbb{R}^m$. For any $\epsilon > 0$, Deville et al. [50] show that there exists a polyhedral set $\mathcal{V}$ (see Theorem 1.1 in [50]) such that

$$\mathcal{V} \subseteq \mathcal{U} \subseteq (1+\epsilon)\cdot\mathcal{V}. \tag{2.7}$$

Note that the number of polyhedral inequalities that describes $\mathcal{V}$ could be exponential in m and $1/\epsilon$. Consider instances of the two-stage adjustable problem (1.1) with random second-stage matrix $\tilde{\boldsymbol{B}}$. Denote β the approximation bound given by Theorem 2.2.1 or Theorem 2.2.3 on the performance of affine policies for polyhedral uncertainty sets. Note that β depends only on the distribution of $\tilde{\boldsymbol{B}}$ and does not depend on the description of the polyhedral uncertainty set. Therefore,

$$z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}},\mathcal{V}) \leq \beta \cdot z_{\mathsf{AR}}(\tilde{\boldsymbol{B}},\mathcal{V}),$$

where we use the notation $z(\tilde{\boldsymbol{B}},\mathcal{V})$ to denote the adjustable or affine problem with random matrix $\tilde{\boldsymbol{B}}$ and uncertainty set $\mathcal{V}$. Combining the above inequality with (2.7), we get

$$z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}},\mathcal{U}) \leq (1+\epsilon)\cdot z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}},\mathcal{V}) \leq \beta(1+\epsilon)\cdot z_{\mathsf{AR}}(\tilde{\boldsymbol{B}},\mathcal{V}) \leq \beta(1+\epsilon)\cdot z_{\mathsf{AR}}(\tilde{\boldsymbol{B}},\mathcal{U}).$$

Since $\epsilon > 0$ could be chosen arbitrary small, then

$$z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}},\mathcal{U}) < \beta \cdot z_{\mathsf{AR}}(\tilde{\boldsymbol{B}},\mathcal{U}).$$

i.e., the same approximation bounds of Theorem 2.2.1 and Theorem 2.2.3 hold as well for general convex uncertainty sets.

Remark 2.2.8. We would like to note that our results extend as well for two-stage robust optimization problems (1.1) where the constraints matrices $\boldsymbol{A}$ and $\tilde{\boldsymbol{B}}$ could possibly have some negative components. In fact, the non-negativity assumption on $\boldsymbol{A}$ could be relaxed without loss of generality since our analysis in the chapter depends only on the second stage matrix $\tilde{\boldsymbol{B}}$. We can relax the non-negativity of $\tilde{\boldsymbol{B}}$ under two assumptions:

1. The affine problem $z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}})$ is feasible.

2. For each row $i \in [m]$ of $\tilde{\boldsymbol{B}}$,

$$\mu_i = \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}[\tilde{B}_{ij}] > 0.$$

In fact, in the proof of Theorem 2.2.1 and Theorem 2.2.3, we did not require the matrix $\tilde{\boldsymbol{B}}$ to be non-negative but we used only the fact that $\mu_i - \tau_i \geq 0$ for small enough τ_i. Hence, our second-stage matrix $\tilde{\boldsymbol{B}}$ could have negative components as long as $\mu_i > 0$ for all rows $i = 1, \ldots, m$. On the other hand, Assumption 1 is required because feasibility of the affine problem is not necessary guaranteed if we relax the non-negativity of both matrices $\boldsymbol{A}$ and $\tilde{\boldsymbol{B}}$.

2.3 Family of worst-case distribution

For any m sufficiently large, the authors in [18] present an instance where affine policy is $\Omega(m^{\frac{1}{2}-\delta})$ away from the optimal adjustable solution. The parameters of the instance in [18] were carefully chosen to achieve the gap $\Omega(m^{\frac{1}{2}-\delta})$. In this section, we show that the family of worst-case instances is not a measure zero set. In fact, we exhibit a distribution and an uncertainty set such that a random instance, $\tilde{\boldsymbol{B}}$ sampled from that distribution achieves a worst-case bound of $\Omega(\sqrt{m})$ with high probability. The coefficients $\tilde{B}_{ij}$ in our bad family of instances are independent but they depend on the dimension of the problem. The instance can be given as follows.

$$
\begin{aligned}
& n = m, \;\; \boldsymbol{A} = \boldsymbol{0}, \;\; \boldsymbol{c} = \boldsymbol{0}, \;\; \boldsymbol{d} = \boldsymbol{e}, \;\; \mathcal{X} = \mathbb{R}^n_+ \\
& \mathcal{U} = \mathsf{conv}\left(\boldsymbol{0}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\right) \;\; \text{where } \boldsymbol{v}_i = \frac{1}{\sqrt{m}}(\boldsymbol{e} - \boldsymbol{e}_i) \; \forall i \in [m]. \\
& \tilde{B}_{ij} = \begin{cases} 1 & \text{if } \; i = j \\ \frac{1}{\sqrt{m}} \cdot \tilde{u}_{ij} & \text{if } \; i \neq j \end{cases} \;\; \text{where for all } i \neq j, \tilde{u}_{ij} \text{ are i.i.d. uniform}[0,1].
\end{aligned}
\tag{2.8}
$$

Theorem 2.3.1. *For the instance defined in* (2.8)*, we have with probability at least* $1 - 1/m$*,*

$$
z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}}) = \Omega(\sqrt{m}) \cdot z_{\mathsf{AR}}(\tilde{\boldsymbol{B}}).
$$

As a byproduct, we also tighten the lower bound on the performance of affine policy to $\Omega(\sqrt{m})$ improving from the lower bound of $\Omega(m^{\frac{1}{2}-\delta})$ in [18]. We would like to note that both uncertainty set and distribution of coefficients in our instance (2.8) are carefully chosen to achieve the worst-case gap. Our analysis suggests that to obtain bad instances for affine policies, we need to generate instances using a structured distribution as above and it may not be easy to obtain bad instances in a completely random setting as observed in extensive empirical studies.

To prove Theorem 2.3.1, we introduce the following lemma which shows a deterministic bad instance where the optimal affine solution is $\Theta(\sqrt{m})$ away from the optimal adjustable solution.

Lemma 2.3.2. *Consider the two-stage adjustable problem* (1.1) *where:*

$$\begin{aligned} & n = m, \;\; \boldsymbol{A} = \boldsymbol{0}, \;\; \boldsymbol{c} = \boldsymbol{0}, \;\; \boldsymbol{d} = \boldsymbol{e}, \;\; \mathcal{X} = \mathbb{R}^n_+ \\ & \mathcal{U} = \mathsf{conv}\,(\boldsymbol{0}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_m) \;\; \textit{where } \boldsymbol{\nu}_i = \frac{1}{\sqrt{m}}(\boldsymbol{e} - \boldsymbol{e}_i) \; \forall i \in [m]. \\ & B_{ij} = \begin{cases} 1 & \textit{if } \; i = j \\ \frac{1}{\sqrt{m}} & \textit{if } \; i \neq j \end{cases} \end{aligned} \tag{2.9}$$

Then, $z_{\mathsf{Aff}}(\boldsymbol{B}) = \Omega(\sqrt{m}) \cdot z_{\mathsf{AR}}(\boldsymbol{B})$.

Proof. First, let us prove that $z_{\mathsf{AR}}(\boldsymbol{B}) \leq 1$. It is sufficient to define an adjustable solution only for the extreme points of $\mathcal{U}$ because the constraints are linear. We define the following solution for all $i \in [m]$,

$$\boldsymbol{x} = \boldsymbol{0}, \qquad \boldsymbol{y}(\boldsymbol{0}) = \boldsymbol{0}, \qquad \boldsymbol{y}(\boldsymbol{e}_i) = \boldsymbol{e}_i, \qquad \boldsymbol{y}(\boldsymbol{\nu}_i) = \frac{1}{m}\boldsymbol{e}.$$

We have $\boldsymbol{B}\boldsymbol{y}(\boldsymbol{0}) = \boldsymbol{0}$. For all $i \in [m]$,

$$\boldsymbol{B}\boldsymbol{y}(\boldsymbol{e}_i) = \boldsymbol{e}_i + \frac{1}{\sqrt{m}}(\boldsymbol{e} - \boldsymbol{e}_i) \geq \boldsymbol{e}_i$$

and

$$\boldsymbol{B}\boldsymbol{y}(\boldsymbol{\nu}_i) = \frac{1}{m}\boldsymbol{B}\boldsymbol{e} = \left(\frac{1}{m} + \frac{m-1}{m\sqrt{m}}\right)\boldsymbol{e} \geq \frac{1}{\sqrt{m}}\boldsymbol{e} \geq \boldsymbol{\nu}_i.$$

Therefore, the solution defined above is feasible. Moreover, the cost of our feasible solution is 1 because for all $i \in [m]$, we have

$$\boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{e}_i) = \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{\nu}_i) = 1.$$

Hence, $z_{\mathsf{AR}}(\boldsymbol{B}) \leq 1$. Now, it is sufficient to prove that $z_{\mathsf{Aff}}(\boldsymbol{B}) = \Omega(\sqrt{m})$. From Lemma 8 in Bertsimas and Goyal [18], since our instance is symmetric, i.e., the uncertainty set $\mathcal{U}$ and the dualized uncertainty set $\mathcal{W}$ are permutation invariant, there exists an optimal solution for the affine problem (1.2) of the following form $\boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}$ for $\boldsymbol{h} \in \mathcal{U}$ where

$$\boldsymbol{P} = \begin{pmatrix} \theta & \mu & \dots & \mu \\ \mu & \theta & \dots & \mu \\ \vdots & \vdots & \ddots & \vdots \\ \mu & \mu & \dots & \theta \end{pmatrix} \tag{2.10}$$

and $\boldsymbol{q} = \lambda \boldsymbol{e}$. We have $\boldsymbol{y}(\boldsymbol{0}) = \lambda \boldsymbol{e} \geq \boldsymbol{0}$ hence

$$\lambda \geq 0. \tag{2.11}$$

We know that

$$z_{\mathsf{Aff}}(\boldsymbol{B}) \geq \boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{0}) = \lambda m. \tag{2.12}$$

Case 1: If $\lambda \geq \frac{1}{6\sqrt{m}}$, then from (2.12) we have $z_{\mathsf{Aff}}(\boldsymbol{B}) \geq \frac{\sqrt{m}}{6}$.

Case 2: If $\lambda \leq \frac{1}{6\sqrt{m}}$. We have,

$$\boldsymbol{y}(\boldsymbol{e}_1) = (\theta + \lambda)\boldsymbol{e}_1 + (\mu + \lambda)(\boldsymbol{e} - \boldsymbol{e}_1).$$

By feasibility of the solution, we have $\boldsymbol{B}\boldsymbol{y}(\boldsymbol{e}_1) \geq \boldsymbol{e}_1$, hence

$$(\theta + \lambda) + \frac{1}{\sqrt{m}}(m-1)(\mu + \lambda) \geq 1.$$

Therefore $\theta + \lambda \geq \frac{1}{2}$ or $\frac{1}{\sqrt{m}}(m-1)(\mu + \lambda) \geq \frac{1}{2}$.

Case 2.1: Suppose $\frac{1}{\sqrt{m}}(m-1)(\mu + \lambda) \geq \frac{1}{2}$. Therefore,

$$z_{\mathsf{Aff}}(\boldsymbol{B}) \geq \boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{e}_1) = \theta + \lambda + (m-1)(\mu + \lambda) \geq \frac{\sqrt{m}}{2}.$$

where the last inequality holds because $\theta + \lambda \geq 0$ as $\boldsymbol{y}(\boldsymbol{e}_1) \geq \boldsymbol{0}$.

Case 2.2: Now suppose we have the other inequality i.e., $\theta + \lambda \geq \frac{1}{2}$. Recall that we have $\lambda \leq \frac{1}{6\sqrt{m}}$ as well. Therefore,

$$\theta \geq \frac{1}{2} - \frac{1}{6\sqrt{m}} \geq \frac{1}{3}.$$

We have,

$$\boldsymbol{y}(\boldsymbol{v}_1) = \frac{1}{\sqrt{m}}\left((\theta + (m-2)\mu)(\boldsymbol{e} - \boldsymbol{e}_1) + (m-1)\mu\boldsymbol{e}_1\right) + \lambda\boldsymbol{e}.$$

Therefore,

$$\begin{aligned} z_{\mathsf{Aff}}(\boldsymbol{B}) \geq \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{v}_1) &= \frac{1}{\sqrt{m}}((m-1)\theta + (m-1)^2\mu) + \lambda m \\ &\geq \frac{m-1}{\sqrt{m}}\left(\frac{1}{3} + (m-1)\mu\right). \end{aligned} \tag{2.13}$$

where the last inequality follows from $\lambda \geq 0$ and $\theta \geq \frac{1}{3}$.

Case 2.2.1: If $\mu \geq 0$ then from (2.13)

$$z_{\mathsf{Aff}}(\boldsymbol{B}) \geq \frac{m-1}{3\sqrt{m}} = \Omega(\sqrt{m}).$$

Case 2.2.2: Now suppose that $\mu < 0$, by non-negativity of $\boldsymbol{y}(\boldsymbol{v}_1)$ we have,

$$\frac{m-1}{\sqrt{m}}\mu + \lambda \geq 0$$

i.e.,

$$\mu \geq \frac{-\lambda\sqrt{m}}{m-1},$$

and from (2.13)

$$
\begin{aligned}
z_{\mathsf{Aff}}(\boldsymbol{B}) &\geq \frac{m-1}{\sqrt{m}}\left(\frac{1}{3}+(m-1)\mu\right)\\
&\geq \frac{m-1}{\sqrt{m}}\left(\frac{1}{3}-\lambda\sqrt{m}\right)\\
&\geq \frac{m-1}{\sqrt{m}}\left(\frac{1}{3}-\frac{1}{6}\right)=\frac{m-1}{6\sqrt{m}}=\Omega(\sqrt{m}).
\end{aligned}
$$

We conclude that in all cases $z_{\mathsf{Aff}}(\boldsymbol{B}) = \Omega(\sqrt{m})$ and consequently $z_{\mathsf{Aff}}(\boldsymbol{B}) = \Omega(\sqrt{m}) \cdot z_{\mathsf{AR}}(\boldsymbol{B})$.

□

Proof of Theorem 2.3.1

Denote

$$\mathcal{W} = \{\boldsymbol{w} \in \mathbb{R}^m_+ \mid \boldsymbol{B}^T\boldsymbol{w} \leq \bar{d}\boldsymbol{e}\}$$

and

$$\tilde{\mathcal{W}} = \{\boldsymbol{w} \in \mathbb{R}^m_+ \mid \tilde{\boldsymbol{B}}^T\boldsymbol{w} \leq \bar{d}\boldsymbol{e}\}$$

where $\boldsymbol{B}$ is defined in (2.9) and $\tilde{\boldsymbol{B}}$ is defined in (2.8). We know for all i, j in $\{1, \ldots, m\}$ that $\tilde{B}_{ij} \leq B_{ij}$. Hence, for any $\boldsymbol{w} \in \mathcal{W}$, we have $\tilde{\boldsymbol{B}}^T\boldsymbol{w} \leq \boldsymbol{B}^T\boldsymbol{w} \leq \bar{d}\boldsymbol{e}$. Therefore $\boldsymbol{w} \in \tilde{\mathcal{W}}$ and consequently $\mathcal{W} \subseteq \tilde{\mathcal{W}}$. Now, suppose $\boldsymbol{w} \in \tilde{\mathcal{W}}$, we have for all $i \in [m]$,

$$w_i + \frac{1}{\sqrt{m}}\sum_{\substack{j=1\\ j\neq i}}^{m} \tilde{u}_{ji}w_j \leq \bar{d}. \tag{2.14}$$

By taking the sum over $i \in [m]$, dividing by m and rearranging, we get

$$\sum_{i=1}^{m} w_i\left(\frac{1}{m}+\frac{1}{m\sqrt{m}}\sum_{\substack{j=1\\ j\neq i}}^{m}\tilde{u}_{ij}\right) \leq \bar{d}. \tag{2.15}$$

We apply Hoeffding's inequality [49] (see appendix B.2) with $\tau = \sqrt{\frac{\log m}{m-1}}$,

$$\mathbb{P}\left(\frac{\sum_{\substack{j=1\\j\neq i}}^m \tilde{u}_{ij}}{m-1} \geq \frac{1}{2} - \tau\right) \geq 1 - \exp\left(-2(m-1)\tau^2\right) = 1 - \frac{1}{m^2},$$

and we take a union bound over $j = 1, \ldots, m$, we get

$$\mathbb{P}\left(\frac{\sum_{\substack{j=1\\j\neq i}}^m \tilde{u}_{ij}}{m-1} \geq \frac{1}{2} - \tau \ \ \forall j = 1, \ldots, m\right) \geq \left(1 - \frac{1}{m^2}\right)^m \geq 1 - \frac{1}{m}, \tag{2.16}$$

where the last inequality follows from Bernoulli's inequality. Therefore, we conclude from (2.15) and (2.16), that with probability at least $1 - \frac{1}{m}$ we have

$$\beta \sum_{i=1}^m w_i \leq \bar{d}$$

where

$$\beta = \frac{1}{m} + \frac{m-1}{m\sqrt{m}}(\frac{1}{2} - \tau) \geq \frac{1}{4\sqrt{m}}$$

for m sufficiently large. Note from (2.14) that for all i we have $w_i \leq \bar{d}$. Hence with probability at least $1 - \frac{1}{m}$, we have for all $i = 1, \ldots, m$

$$\boldsymbol{B}_i^T \boldsymbol{w} = w_i + \frac{1}{\sqrt{m}} \sum_{\substack{j=1\\j\neq i}}^m w_j \leq \bar{d} + \frac{\bar{d}}{\beta\sqrt{m}} \leq 5 \cdot \bar{d}.$$

Therefore, $\boldsymbol{w} \in 5 \cdot \mathcal{W}$ for any $\boldsymbol{w}$ in $\mathcal{W}$ and consequently we have with probability at least $1 - \frac{1}{m}$ that, $\tilde{\mathcal{W}} \subseteq 5 \cdot \mathcal{W}$. All together we have proved with probability at least $1 - \frac{1}{m}$, that

$$\mathcal{W} \subseteq \tilde{\mathcal{W}} \subseteq 5 \cdot \mathcal{W}.$$

This implies with probability at least $1 - \frac{1}{m}$, that $z_{\mathsf{d-Aff}}(\tilde{\boldsymbol{B}}) \geq z_{\mathsf{d-Aff}}(\boldsymbol{B})$ and $z_{\mathsf{d-AR}}(\boldsymbol{B}) \geq \frac{z_{\mathsf{d-AR}}(\tilde{\boldsymbol{B}})}{5}$.

We know from from Lemma 2.2.5 and Lemma 2.2.4 that the dualized and primal are the same both for the adjustable problem and affine problem. Hence, with probability at least $1 - \frac{1}{m}$, we have $z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}}) \geq z_{\mathsf{Aff}}(\boldsymbol{B})$ and $z_{\mathsf{AR}}(\boldsymbol{B}) \geq \frac{z_{\mathsf{AR}}(\tilde{\boldsymbol{B}})}{5}$.

Moreover, we know from Lemma 2.3.2 that $z_{\mathsf{Aff}}(\boldsymbol{B}) \geq \Omega(\sqrt{m}) \cdot z_{\mathsf{AR}}(\boldsymbol{B})$. Therefore, with probability at least $1 - \frac{1}{m}$,

$$z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}}) \geq \Omega(\sqrt{m}) z_{\mathsf{AR}}(\tilde{\boldsymbol{B}}).$$

2.4 Performance of affine policy: Empirical study

In this section, we present a computational study to test the empirical performance of affine policy for the two-stage adjustable problem (1.1) on random instances.

Experimental setup. We consider two classes of distributions for generating random instances: i) Coefficients of $\tilde{\boldsymbol{B}}$ are i.i.d. uniform $[0,1]$, and ii) Coefficients of $\tilde{\boldsymbol{B}}$ are absolute value of i.i.d. standard Gaussian. We consider the following *budget of uncertainty* set.

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} h_i \leq \sqrt{m} \right\}. \tag{2.17}$$

Note that the set (2.17) is widely used in both theory and practice and arises naturally as a consequence of concentration of sum of independent uncertain demand requirements. We would like to also note that the adjustable problem over this budget of uncertainty, $\mathcal{U}$ is hard to approximate within a factor better than $O(\frac{\log n}{\log \log n})$ [24]. We consider $n = m, \boldsymbol{d} = \boldsymbol{e}$. Also, we consider $\boldsymbol{c} = \boldsymbol{0}, \boldsymbol{A} = \boldsymbol{0}$. We restrict to this case in order to compute the optimal adjustable solution in a reasonable time by solving a single MIP. For the general problem, computing the optimal adjustable solution requires solving a sequence of MIPs each one of which is significantly challenging to solve. We would like to note though that our analysis does not depend on the first stage cost $\boldsymbol{c}$ and matrix $\boldsymbol{A}$ and affine policy can be computed efficiently even without this assumption. We consider values of m from 10 to 50 and consider 20 instances for each value of m. We report the ratio $r = z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}})/z_{\mathsf{AR}}(\tilde{\boldsymbol{B}})$ in Table 2.1. In particular, for each value of m, we report the average ratio

r_{avg}, the maximum ratio r_{max}, the running time of adjustable policy $T_{\mathsf{AR}}(s)$ and the running time of affine policy $T_{\mathsf{Aff}}(s)$. We first give a compact LP formulation for the affine problem (1.2) and a compact MIP formulation for the separation of the adjustable problem(1.1).

LP formulations for the affine policies. The affine problem (1.2) can be reformulated as follows

$$
\begin{aligned}
z_{\mathsf{Aff}}(\boldsymbol{B}) = \min_{\boldsymbol{x}}\ & \boldsymbol{c}^T\boldsymbol{x} + z \\
& z \geq \boldsymbol{d}^T\left(\boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}\right) \quad \forall \boldsymbol{h} \in \mathcal{U} \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\left(\boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}\right) \ \geq\ \boldsymbol{h} \quad \forall \boldsymbol{h} \in \mathcal{U} \\
& \boldsymbol{P}\boldsymbol{h} + \boldsymbol{q} \ \geq\ \boldsymbol{0} \quad \forall \boldsymbol{h} \in \mathcal{U} \\
& \boldsymbol{x} \in \mathcal{X}.
\end{aligned}
$$

Note that this formulation has infinitely many constraints but we can write a compact LP formulation using standard techniques from duality. The LP formulation is given in Lemma A.0.2 in Appendix A.

MIP Formulation for the adjustable problem (1.1). For the adjustable problem (1.1), we show that the separation problem (2.18) can be formulated as a mixed integer program (MIP). The separation problem can be formulated as follows: Given $\hat{\boldsymbol{x}}$ and $\hat{z}$ decide whether

$$
\max\left\{(\boldsymbol{h} - \boldsymbol{A}\hat{\boldsymbol{x}})^T\boldsymbol{w} \mid \boldsymbol{w} \in \mathcal{W}, \boldsymbol{h} \in \mathcal{U}\right\} > \hat{z} \tag{2.18}
$$

The correctness of formulation (2.18) follows from equation (B.1) in the proof of Lemma 2.2.4 in Appendix B.1. The constraints in (2.18) are linear but the objective function contains a bilinear term, $\boldsymbol{h}^T\boldsymbol{w}$. We linearize this using a standard *digitized reformulation*. In particular, we consider finite bit representations of continuous variables, h_i nd w_i to desired accuracy and introduce additional binary variables, α_{ik}, β_{ik} where α_{ik} and β_{ik} represents the k^{th} bits of h_i and w_i respectively. Now, for any $i \in [m]$, $h_i \cdot w_i$ can be expressed as a bilinear expression with products of binary variables, $\alpha_{ik} \cdot \beta_{ij}$ which can be linearized using additional variable γ_{ijk} and standard linear inequalities: $\gamma_{ijk} \leq \beta_{ij}$, $\gamma_{ijk} \leq \alpha_{ik}$, $\gamma_{ijk} + 1 \geq \alpha_{ik} + \beta_{ij}$. The complete MIP formulation and the proof of correctness is presented in Appendix B.3.

For general $\boldsymbol{A} \neq 0$, we need to solve a sequence of MIPs to find the optimal adjustable solution. In order to compute the optimal adjustable solution in a reasonable time, we assume $\boldsymbol{A} = 0, \boldsymbol{c} = 0$ in our experimental setting so that we only need to solve one MIP.

Results. In our experiments, we observe that the empirical performance of affine policy is near-optimal. In particular, the performance is significantly better than the theoretical performance bounds implied in Theorem 2.2.1 and Theorem 2.2.3. For instance, Theorem 2.2.1 implies that affine policy is a 2-approximation with high probability for i.i.d. random instances from a uniform distribution (see Corollary 2.2.2). However, in our experiments, we observe that the optimality gap for affine policies is at most 4% (i.e. approximation ratio of at most 1.04). The same observation holds for Gaussian distributions as well Theorem 2.2.3 gives an approximation bound of $O(\sqrt{\log mn})$. We would like to remark that we are not able to report the ratio r for large values of m because the adjustable problem is computationally very challenging and for $m \geq 40$, MIP does not solve within a time limit of 3 hours for most instances . On the other hand, affine policy scales very well and the average running time is few seconds even for large values of m. This demonstrates the power of affine policies that can be computed efficiently and give good approximations for a large class of instances.

m	r_{avg}	r_{max}	$T_{\mathsf{AR}}(s)$	$T_{\mathsf{Aff}}(s)$
10	1.01	1.03	10.55	0.01
20	1.02	1.04	110.57	0.23
30	1.01	1.02	761.21	1.29
50	**	**	**	14.92

(a) Uniform

m	r_{avg}	r_{max}	$T_{\mathsf{AR}}(s)$	$T_{\mathsf{Aff}}(s)$
10	1.00	1.03	12.95	0.01
20	1.01	1.03	217.08	0.39
30	1.01	1.03	594.15	1.15
50	**	**	**	13.87

(b) Folded Normal

Table 2.1: Comparison on the performance and computation time of affine policy and optimal adjustable policy for uniform and folded normal distributions. For 20 instances, we compute $z_{\mathsf{Aff}}(\tilde{\boldsymbol{B}})/z_{\mathsf{AR}}(\tilde{\boldsymbol{B}})$ and present the average and max ratios. Here, $T_{\mathsf{AR}}(s)$ denotes the running time for the adjustable policy and $T_{\mathsf{Aff}}(s)$ denotes the running time for affine policy in seconds. ** Denotes the cases when we set a time limit of 3 hours. These results are obtained using Gurobi 7.0.2 on a 16-core server with 2.93GHz processor and 56GB RAM.

Chapter 3: Affine policies for budget of uncertainty sets

3.1 Introduction.

In this chapter, we study the performance of affine policies for two-stage adjustable robust optimization problem with fixed recourse and uncertain right hand side belonging to a budgeted uncertainty set. This is an important class of uncertainty sets widely used in practice where we can specify a budget on the adversarial deviations of the uncertain parameters from the nominal values to adjust the level of conservatism. The two-stage adjustable robust optimization problem is hard to approximate within a factor better than $\Omega(\frac{\log n}{\log \log n})$ even for budget of uncertainty sets where n is the number of decision variables. Affine policies, where the second-stage decisions are constrained to be an affine function of the uncertain parameters, provide a tractable approximation for the problem and have been observed to exhibit good empirical performance. We show that affine policies give an $O(\frac{\log n}{\log \log n})$-approximation for the two-stage adjustable robust problem with fixed non-negative recourse for budgeted uncertainty sets. This matches the hardness of approximation and therefore, surprisingly affine policies provide an optimal approximation for the problem (up to a constant factor). We also show strong theoretical performance bounds for affine policy for significantly more general class of intersection of budgeted sets including disjoint constrained budgeted sets, permutation invariant sets and general intersection of budgeted sets. Our analysis relies on showing the existence of a near-optimal feasible affine policy that satisfies certain nice structural properties. Based on these structural properties, we also present an alternate algorithm to compute near-optimal affine solution that is significantly faster than computing the optimal affine policy by solving a large linear program.

The rest of this chapter is organized as follows. In Section 3.2, we present our performance analysis for affine policies on budget of uncertainty sets. Then, we focus on the analysis of a more general class of uncertainty sets including disjoint constrained budgeted sets (Section 3.3) and general intersection of budgeted sets (Section 3.4). In Section 3.5, we present our new faster algorithm to compute affine policies. Finally, we discuss the performance of affine policies for a broader class of two-stage robust problem where we relax the assumptions on the constraints matrices. In particular, we discuss the case of general matrix $\boldsymbol{B}$ in Section 3.6 and the case of uncertainty in constraint matrix $\boldsymbol{A}$ in Section 3.7.

Notations. For simplicity, we refer in this chapter to $z_{\mathsf{AR}}(\boldsymbol{c},\boldsymbol{d},\boldsymbol{A},\boldsymbol{B},\mathcal{U})$ as $z_{\mathsf{AR}}(\mathcal{U})$ and to $z_{\mathsf{Aff}}(\boldsymbol{c},\boldsymbol{d},\boldsymbol{A},\boldsymbol{B},\mathcal{U})$ as $z_{\mathsf{Aff}}(\mathcal{U})$. We also assume that $\mathcal{U} \subseteq [0,1]^m$ and $\forall i \in [m], \boldsymbol{e}_i \in \mathcal{U}$. This assumption is without loss of generality since we can scale the constraint matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ to satisfy the assumption without changing the optimal.

3.2 Performance analysis for budget of uncertainty sets.

In this section, we consider the class of budget of uncertainty sets (1.5) given by

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} w_i h_i \leq 1 \right\}.$$

As we mention earlier, this class is widely used in the literature of robust optimization both in theory and practice. It provides the flexibility to adjust the level of conservatism in terms of probabilistic bounds on constraint violations. A special case of this class is when w_i are all equal to $\frac{1}{k}$ for some parameter $k \in \mathbb{N}$. In particular, in this case we have

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} h_i \leq k \right\}. \tag{3.1}$$

The parameter k is the *budget of uncertainty* that controls the conservatism of the uncertainty model. This special class (3.1) of budgeted sets is also known as the *cardinality constrained set* or *k-ones polytope*. Recall the two-stage adjustable problem

$$
\begin{aligned}
z_{\mathsf{AR}}(\mathcal{U}) = \min_{\boldsymbol{x}} \; & \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \min_{\boldsymbol{y}(\boldsymbol{h})} \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{h}) \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) \;\geq\; \boldsymbol{h}, \quad \forall \boldsymbol{h} \in \mathcal{U} \\
& \boldsymbol{x} \in \mathcal{X} \\
& \boldsymbol{y}(\boldsymbol{h}) \in \mathbb{R}^n_+, \quad \forall \boldsymbol{h} \in \mathcal{U}.
\end{aligned}
$$

The two-stage adjustable problem (1.1) is known to be $\Omega(\frac{\log n}{\log \log n})$-hard to approximate under the class of budget of uncertainty sets even in the special case of (3.1) (Feige et al. [24]). We show that surprisingly the performance bound for affine policy matches this hardness of approximation. In particular, we show that affine policy gives $O(\frac{\log n}{\log \log n})$-approximation for (1.1) under budget of uncertainty sets (1.5).

Theorem 3.2.1. *Consider the two-stage adjustable problem* (1.1) *where $\mathcal{U}$ is the budget of uncertainty set* (1.5) *. Then,*

$$
z_{\mathsf{Aff}}(\mathcal{U}) = O\left(\frac{\log n}{\log \log n}\right) \cdot z_{\mathsf{AR}}(\mathcal{U}).
$$

Our analysis significantly improves over the previous best known bound of $O(\sqrt{m})$ for the performance of affine policies for budget of uncertainty sets. In fact, Bertsimas and Goyal [18] shows that affine policy gives $O(\sqrt{m})$-approximation to the adjustable problem (1.1) under this class of uncertainty sets and $\mathcal{X} = \mathbb{R}^n_+$. Bertsimas and Bidkhori [36] provide a geometric bound $O(\frac{k^2+mk}{k^2+m})$ in the special case of (1.5) where all $w_i = 1/k$ and $\mathcal{X} = \mathbb{R}^n_+$. This bound is also $O(\sqrt{m})$ in the worst-case for $k = \sqrt{m}$.

The above two-stage adjustable robust problem (1.1) has also been considered in the context of combinatorial optimization problems such as network design under demand uncertainty where the constraint matrices, $\boldsymbol{A}, \boldsymbol{B} \in \{0,1\}^{m\times n}$ and the first-stage and second-stage decisions are constrained to be binary (see for instance, Dhamdhere et al. [19], Feige et al. [24], Gupta et al. [21] and [51]). Feige et al. [24] and Gupta et al. [21] give an $O(\log n)$-approximation for (1.1) for the special case when $\boldsymbol{A} = \boldsymbol{B} \in \{0,1\}^{m\times n}$, first and second-stage costs are proportional, i.e., $\boldsymbol{d} = \lambda \cdot \boldsymbol{c}$ for some constant $\lambda \geq 1$ and a budget of uncertainty set with $w_i = 1/k$. However, we would like to note that the focus of this stream of work is to design approximation algorithms for combinatorial optimization problems where the decisions are constrained to be binary. Moreover, the algorithms are not restricted to and do not necessarily give decision rules or functional policy approximations for the two-stage problem. In contrast, the focus of our work is to analyze the performance of affine policies for the two-stage adjustable robust problem (1.1) that are widely used in practice and exhibit strong empirical performance.

Since our performance bound in Theorem 3.2.1 matches the hardness of approximation, affine policy provides an optimal approximation for (1.1) for budget of uncertainty sets. In particular, there is no polynomial time algorithm whose worst-case approximation is better than affine policies by more than a constant factor. Note that the above statements relate to the worst-case performance. For particular instances, it may be possible to get better solutions than affine policies.

3.2.1 Construction of our affine solution.

Our analysis is based on constructing a *good* approximate affine solution that has a worst-case cost being $O\left(\frac{\log n}{\log\log n}\right)$ times the optimal cost $z_{\mathsf{AR}}(\mathcal{U})$. In this section, we present the construction of our affine solution and consequently prove Theorem 3.2.1. Let us first introduce the following notations.

Notations. We consider an optimal solution $\boldsymbol{x}^*, \boldsymbol{y}^*(\boldsymbol{h})$ where $\boldsymbol{h} \in \mathcal{U}$, for the adjustable problem (1.1). Let OPT be the optimal cost for (1.1) and $\text{OPT}_1, \text{OPT}_2$ respectively the first stage cost and the second stage cost associated with $\boldsymbol{x}^*, \boldsymbol{y}^*(\boldsymbol{h})$, i.e.

$$\begin{aligned}
\text{OPT}_1 &= \boldsymbol{c}^T \boldsymbol{x}^* \\
\text{OPT}_2 &= \max_{\boldsymbol{h} \in \mathcal{U}} \ \boldsymbol{d}^T \boldsymbol{y}^*(\boldsymbol{h}) \\
\text{OPT} &= \text{OPT}_1 + \text{OPT}_2 = z_{\mathsf{AR}}(\mathcal{U}).
\end{aligned}$$

We would like to remark that this split is not unique since there might be other optimal solutions for (1.1). For all $i \in [m]$, we denote

$$\alpha_i = 1 - (\boldsymbol{A}\boldsymbol{x}^*)_i$$

In particular, if α_i is negative the first stage solution $\boldsymbol{x}^*$ covers the full unit requirement in the i-th component, if not α_i corresponds to the remaining requirement that needs to be covered eventually by the second-stage solution $\boldsymbol{y}^*(\cdot)$.

We denote $z(\boldsymbol{h})$, the cost of covering the requirement $\boldsymbol{h}$ in the second-stage, i.e.

$$z(\boldsymbol{h}) = \min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{h} \right\}. \tag{3.2}$$

We refer to problem (3.2) as the *fractional covering problem*. For any $\mathcal{W} \subseteq [m]$, we denote $\mathbb{1}(\mathcal{W}) \in \mathbb{R}^m$ the indicator of $\mathcal{W}$, i.e.

$$\mathbb{1}(\mathcal{W}) = \sum_{i \in \mathcal{W}} \boldsymbol{e}_i.$$

For simplicity, we use the following notation

$$z(\mathbb{1}(\mathcal{W})) = z(\mathcal{W}).$$

Our construction. For all $i \in [m]$, recall $z(\boldsymbol{e}_i)$ the optimal cost to cover component $\boldsymbol{e}_i$ in the second stage as defined in (3.2). Let $\boldsymbol{v}_i$ be the optimal corresponding solution, i.e.,

$$\boldsymbol{v}_i \in \arg\min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{e}_i \right\}.$$

We split the components $\{1, 2, \ldots, m\}$ based on a threshold into two sets $\mathcal{I}$ and its complement $\mathcal{I}^c$:

$$\mathcal{I} = \left\{ i \in [m] \;\middle|\; \alpha_i > 0 \text{ and } \frac{\alpha_i z(\boldsymbol{e}_i)}{w_i} \leq \beta \cdot \text{OPT} \right\}$$
$$\mathcal{I}^c = [m] \setminus \mathcal{I},$$

where

$$\beta = \frac{4 \log n}{\log \log n}.$$

We cover a fraction of $\mathcal{I}$ (*inexpensive components*) using a linear solution and the remaining fraction of $\mathcal{I}$ along with $\mathcal{I}^c$ (*expensive components*) using a static solution.

Linear part. We cover a fraction of the components of $\mathcal{I}$ using the following linear solution for any $\boldsymbol{h} \in \mathcal{U}$,

$$\boldsymbol{y}_{\mathsf{Lin}}(\boldsymbol{h}) = \sum_{i \in \mathcal{I}} \alpha_i h_i \boldsymbol{v}_i. \tag{3.3}$$

Static part. We use a static solution to cover the remaining components $\boldsymbol{e}_i$ where $i \in \mathcal{I}^c$ and $(1-\alpha_i)^+ \boldsymbol{e}_i$ for $i \in \mathcal{I}$. In particular, we consider the following static problem

$$(\boldsymbol{x}_{\mathsf{Sta}}, \boldsymbol{y}_{\mathsf{Sta}}) \in \underset{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \geq \boldsymbol{0}}{\arg\min} \left\{ \boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \sum_{i \in \mathcal{I}^c} \boldsymbol{e}_i + \sum_{i \in \mathcal{I}} (1-\alpha_i)^+ e_i \right\}, \tag{3.4}$$

and denote

$$z_{\mathsf{Sta}} = \boldsymbol{c}^T \boldsymbol{x}_{\mathsf{Sta}} + \boldsymbol{d}^T \boldsymbol{y}_{\mathsf{Sta}}.$$

Therefore our candidate affine solution is

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{x}_{\mathsf{Sta}} \\ \boldsymbol{y}(\boldsymbol{h}) &= \boldsymbol{y}_{\mathsf{Lin}}(\boldsymbol{h}) + \boldsymbol{y}_{\mathsf{Sta}}, \ \forall \boldsymbol{h} \in \mathcal{U}. \end{aligned} \tag{3.5}$$

Feasibility. We first show that our candidate solution (3.5) is feasible for the adjustable problem (1.1). The proof is a direct consequence of our construction. In particular, we have the following lemma.

Lemma 3.2.2. *The affine solution in* (3.5) *is feasible for the adjustable problem* (1.1).

Proof. We have,

$$\boldsymbol{B}\boldsymbol{y}_{\mathsf{Lin}}(\boldsymbol{h}) = \sum_{i \in \mathcal{I}} \alpha_i h_i \boldsymbol{B}\boldsymbol{v}_i \geq \sum_{i \in \mathcal{I}} \alpha_i h_i \boldsymbol{e}_i,$$

and

$$\boldsymbol{A}\boldsymbol{x}_{\mathsf{Sta}} + \boldsymbol{B}\boldsymbol{y}_{\mathsf{Sta}} \geq \sum_{i \in \mathcal{I}^c} \boldsymbol{e}_i + \sum_{i \in \mathcal{I}} (1-\alpha_i)^+ e_i \geq \sum_{i \in \mathcal{I}^c} h_i \boldsymbol{e}_i + \sum_{i \in \mathcal{I}} h_i (1-\alpha_i)^+ e_i$$

where the last inequality holds because $h_i \in [0,1]$ for all $i \in [m]$. Therefore, the solution in (3.5) verifies

$$\begin{aligned} &\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) \geq \sum_{i \in \mathcal{I}^c} h_i \boldsymbol{e}_i + \sum_{i \in \mathcal{I}} ((1-\alpha_i)^+ + \alpha_i) h_i \boldsymbol{e}_i \geq \boldsymbol{h}. \\ &\boldsymbol{x} \in \mathcal{X} \\ &\boldsymbol{y}(\boldsymbol{h}) \geq \boldsymbol{0}, \quad \forall \boldsymbol{h} \in \mathcal{U}. \end{aligned}$$

□

We would like to remark that the construction of the linear ans static parts not only depends on the uncertainty set $\mathcal{U}$, but also depends on all the parameters of the instance, i.e., $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{c}, \boldsymbol{d}$. This is in contrast to the analysis in [18] where the construction of affine policies depends only on $\mathcal{U}$.

3.2.2 Performance analysis.

We analyze separately the cost of the static and linear parts. For the linear part, the cost analysis is a direct consequence of our construction. In fact, we leave only inexpensive scenarios to the linear part, i.e., scenarios $\alpha_i \boldsymbol{e}_i$ such that $\alpha_i z(\boldsymbol{e}_i)/w_i$ is less than the threshold $\beta \cdot \text{OPT}$. We know that for all $\boldsymbol{h} \in \mathcal{U}$, we have $\sum_{i=1}^m w_i h_i \leq 1$. Hence, the cost of linear part is at most $\beta \cdot \text{OPT}$. In particular, we have the following lemma.

Lemma 3.2.3 (Cost of Linear part). *The cost of the linear part $\boldsymbol{y}_{\mathsf{Lin}}(\boldsymbol{h})$ defined in* (3.3) *is at most $\beta \cdot \text{OPT}$ for any $\boldsymbol{h} \in \mathcal{U}$.*

Proof. We have for all $\boldsymbol{h} \in \mathcal{U}$,

$$\boldsymbol{d}^T \boldsymbol{y}_{\mathsf{Lin}}(\boldsymbol{h}) = \sum_{i \in \mathcal{I}} \alpha_i h_i \boldsymbol{d}^T \boldsymbol{v}_i = \sum_{i \in \mathcal{I}} \alpha_i h_i z(\boldsymbol{e}_i) \leq \beta \cdot \text{OPT} \cdot \sum_{i \in \mathcal{I}} w_i h_i \leq \beta \cdot \text{OPT},$$

where the first inequality holds because $\alpha_i z(\boldsymbol{e}_i) \leq w_i \beta \cdot \text{OPT}$ for all $i \in \mathcal{I}$ and the second inequality follows as $\sum_{i \in \mathcal{I}} w_i h_i \leq 1$ for all $\boldsymbol{h} \in \mathcal{U}$.

□

The key part is to analyze the cost of the static part. In fact, we show that the cost of the static part is also $O(\beta) \cdot \text{OPT}$. This relies on a structural result on fractional covering problems. Intuitively, let us explain the structural result in the special case (3.1) of the budget of uncertainty set and α_i are 0 or 1. We show that if the cost of covering every single component $\boldsymbol{e}_i$, for $i \in \mathcal{J}$ is *expensive*, i.e. $z(\boldsymbol{e}_i) > \beta \cdot \text{OPT}/k$ and the cost of covering any k components is *inexpensive*, i.e. less than 2OPT. Then, the cost of covering all components of $\mathcal{J}$ is not too costly and can not exceed $\beta \cdot \text{OPT}$. The formal general statement is given in the following lemma.

Lemma 3.2.4 (Structural Result). *Consider* $\boldsymbol{B} \in \mathbb{R}_+^{m \times n}$, $\boldsymbol{d} \in \mathbb{R}_+^n$ *and* $\mathcal{J} \subseteq [m]$. *Let* $z(\boldsymbol{h})$ *be the cost of covering* $\boldsymbol{h}$ *as defined in* (3.2). *Suppose there exists* $\gamma > 0$ *and* $w_i \in (0, 1]$, $\forall i \in \mathcal{J}$ *such that the following two conditions are satisfied:*

1. *for all* $i \in \mathcal{J}$,
$$\frac{z(\boldsymbol{e}_i)}{w_i} > 4\gamma \cdot \frac{\log n}{\log \log n},$$
2. *for all* $\mathcal{W} \subseteq \mathcal{J}$,
$$\sum_{i \in \mathcal{W}} w_i \leq 1 \ \text{ implies } \ z(\mathcal{W}) \leq \gamma.$$

Then,
$$z(\mathcal{J}) \leq 4\gamma \cdot \frac{\log n}{\log \log n}.$$

Lemma 3.2.4 is a generalization of the result in Gupta et al. [21] for the set covering problem (see Theorem 7.1 in [21]). In particular, our result hold for any constraint matrix $\boldsymbol{B}$ and a budgeted set with general w_i for $i \in [m]$, whereas the Gupta et al. [22], discuss the special case when $\boldsymbol{B} \in \{0, 1\}^{m \times n}$, and a budget of uncertainty set with $w_i = 1/k$. Furthermore, we improve the approximation bound from $O(\log n)$ in [21] to $O(\log n / \log \log n)$. We present the proof of Lemma 3.2.4 later in Section 3.2.3. But let us first use the structural result to show that the cost of the static part is $O(\beta) \cdot \mathrm{OPT}$ and consequently prove Theorem 3.2.1. In particular, we have the following lemma.

Lemma 3.2.5 (Cost of Static part). *The cost z_{Sta} of the static part $(\boldsymbol{x}_{\mathsf{Sta}}, \boldsymbol{y}_{\mathsf{Sta}})$ defined in* (3.4) *is* $O(\beta) \cdot \mathrm{OPT}$.

Proof. Consider the following sets

$$\mathcal{J}_1 = \left\{ i \in [m] \;\middle|\; \alpha_i \leq 0 \right\}.$$

and

$$\mathcal{J}_2 = \mathcal{I}^c \setminus \mathcal{J}_1 = \left\{ i \in [m] \;\middle|\; \alpha_i > 0 \text{ and } \frac{\alpha_i z(\boldsymbol{e}_i)}{w_i} > \beta \cdot \mathrm{OPT} \right\}.$$

For $i = 1, \ldots, m$, denote $\boldsymbol{B}_i^T$ the i-th row of $\boldsymbol{B}$ and let $\tilde{\boldsymbol{B}}_i^T = \boldsymbol{B}_i^T / \alpha_i$. We have for $i \in [m]$,

$$\alpha_i z(\boldsymbol{e}_i) = \min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \tilde{\boldsymbol{B}} \boldsymbol{y} \geq \boldsymbol{e}_i \right\}.$$

We apply the structural Lemma 3.2.4 with the parameters $\tilde{\boldsymbol{B}}, \boldsymbol{d}, \mathcal{J}_2$ and $\gamma = \mathrm{OPT}$. Let us verify the assumptions of Lemma 3.2.4. For all $i \in \mathcal{J}_2$, we have

$$\frac{\alpha_i z(\boldsymbol{e}_i)}{w_i} > \beta \cdot \mathrm{OPT} = 4\gamma \cdot \frac{\log n}{\log \log n}.$$

For any $\mathcal{W} \subseteq \mathcal{J}_2$ such that $\sum_{i \in \mathcal{W}} w_i \leq 1$, we have $\boldsymbol{h} = \mathbb{1}(\mathcal{W}) \in \boldsymbol{\mathcal{U}}$. By feasibility of the optimal solution, we know that

$$\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{h}) \geq \boldsymbol{h},$$

which implies,

$$\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{h}) \geq \sum_{i \in \mathcal{W}} \boldsymbol{e}_i - \sum_{i=1}^{m} (1 - \alpha_i) \boldsymbol{e}_i = \sum_{i \in \mathcal{W}} \alpha_i \boldsymbol{e}_i + \sum_{i \notin \mathcal{W}} (\alpha_i - 1) \boldsymbol{e}_i$$

In particular, we have for all $i \in \mathcal{W}$, $(\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{h}))_i \geq \alpha_i$. Moreover, $\boldsymbol{B}$ and $\boldsymbol{y}^*$ are non-negative which implies that

$$\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{h}) \geq \sum_{i\in\mathcal{W}} \alpha_i \boldsymbol{e}_i,$$

and therefore,

$$\tilde{\boldsymbol{B}}\boldsymbol{y}^*(\boldsymbol{h}) \geq \sum_{i\in\mathcal{W}} \boldsymbol{e}_i = \boldsymbol{h}.$$

This means that $\boldsymbol{y}^*(\boldsymbol{h})$ is a feasible solution for the covering problem (3.2) with constraint matrix $\tilde{\boldsymbol{B}}$ and requirement $\boldsymbol{h}$. Therefore,

$$\min_{\boldsymbol{y}\geq\boldsymbol{0}} \left\{ \boldsymbol{d}^T\boldsymbol{y} \;\middle|\; \tilde{\boldsymbol{B}}\boldsymbol{y} \geq \boldsymbol{h} \right\} \leq \boldsymbol{d}^T\boldsymbol{y}^*(\boldsymbol{h}) \leq \mathrm{OPT}_2 \leq \mathrm{OPT} = \gamma.$$

This verifies the second assumption of Lemma 3.2.4. Therefore, from the structural result, we have

$$\min_{\boldsymbol{y}\geq\boldsymbol{0}} \left\{ \boldsymbol{d}^T\boldsymbol{y} \;\middle|\; \tilde{\boldsymbol{B}}\boldsymbol{y} \geq \sum_{i\in\mathcal{J}_2} \boldsymbol{e}_i \right\} \leq 4\gamma\frac{\log n}{\log\log n} = \beta\cdot\mathrm{OPT}.$$

Denote $\boldsymbol{y}_2$ an optimal solution corresponding to the above minimization problem. In particular, we have $\boldsymbol{d}^T\boldsymbol{y}_2 \leq \beta\cdot\mathrm{OPT}$ and

$$\boldsymbol{B}\boldsymbol{y}_2 \geq \sum_{i\in\mathcal{J}_2} \alpha_i \boldsymbol{e}_i.$$

Furthermore, by feasibility of the optimal solution for (1.1), we have

$$\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{0}) \geq \boldsymbol{0},$$

and we know that $\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{0}) \geq \boldsymbol{0}$. This implies,

$$\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{0}) \geq \sum_{i=1}^{m} (\alpha_i - 1)^+ \boldsymbol{e}_i.$$

Putting all together, we have

$$
\begin{aligned}
\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{0}) + \boldsymbol{B}\boldsymbol{y}_2 &\geq \sum_{i=1}^{m}(1-\alpha_i)\boldsymbol{e}_i + \sum_{i=1}^{m}(\alpha_i - 1)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_2}\alpha_i\boldsymbol{e}_i \\
&= \sum_{i=1}^{m}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_2}\alpha_i\boldsymbol{e}_i \\
&= \sum_{i\in\mathcal{I}}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_1}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_2}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_2}\alpha_i\boldsymbol{e}_i \\
&\geq \sum_{i\in\mathcal{I}}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_1}\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_2}\boldsymbol{e}_i \\
&= \sum_{i\in\mathcal{I}}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{I}^c}\boldsymbol{e}_i
\end{aligned}
$$

where the last inequality holds because $\alpha_i \leq 0$ for all $i \in \mathcal{J}_1$ and $(1-\alpha_i)^+ + \alpha_i \geq 1$ for all $i \in \mathcal{J}_2$. Moreover, $\boldsymbol{x}^* \in \mathcal{X}$ and $(\boldsymbol{y}^*(\boldsymbol{0}) + \boldsymbol{y}_2)$ is non-negative. Hence, we have $(\boldsymbol{x}^*, \boldsymbol{y}^*(\boldsymbol{0}) + \boldsymbol{y}_2)$ is a feasible solution for the static problem in (3.4), therefore

$$
z_{\mathsf{Sta}} \leq \boldsymbol{c}^T\boldsymbol{x}^* + \boldsymbol{d}^T y^*(\boldsymbol{0}) + \boldsymbol{d}^T\boldsymbol{y}_2 \leq \text{OPT} + \beta \cdot \text{OPT} = O(\beta)\cdot\text{OPT}.
$$

□

Proof of Theorem 3.2.1. Lemma 3.2.2 show that our affine solution (3.5) is feasible for the adjustable problem (1.1). Lemma 3.2.3 and 3.2.5 show that the cost of the affine solution (3.5) is less than $\beta\cdot\text{OPT} + O(\beta)\cdot\text{OPT} = O(\beta)\cdot\text{OPT}$ which implies that

$$
z_{\mathsf{Aff}}(\mathcal{U}) = O\left(\frac{\log n}{\log\log n}\right)\cdot z_{\mathsf{AR}}(\mathcal{U}).
$$

3.2.3 Proof of the Structural Result.

We give a proof by contradiction. The assumptions in Lemma 3.2.4 can be interpreted as follows. Let $\eta = \frac{4\log n}{\log\log n}$. The first assumption states that the cost of covering any single component, $\boldsymbol{e}_i$ for $i \in \mathcal{J}$ is large (at least $w_i\eta\cdot\gamma$). The second assumption states that the cost of any feasible (integral) scenario, $\mathcal{W} \subseteq \mathcal{J}$ with $\sum_{i \in \mathcal{W}} w_i \leq 1$ is at most γ. We need to show that the cost of covering all the components, $\mathcal{J}$ is at most $\eta \cdot \gamma$. For the sake of contradiction, suppose that $z(\mathcal{J}) > \eta\gamma$. We will construct a feasible scenario, $\mathcal{W} \subseteq \mathcal{J}$ (where $h_i = 1$ for all $i \in \mathcal{W}$ and $\sum_{i \in \mathcal{W}} w_i \leq 1$) where the cost, $z(\mathcal{W}) > \gamma$ violating the second assumption in Lemma 3.2.4. To construct this scenario, we consider the dual of the primal covering problem $z(\mathcal{J})$. The dual is a packing problem where the ratio of the right hand sides and the constraint coefficients is *large* (from the first assumption in the lemma). This allows us to construct an approximate *integral* dual solution of the problem (using randomized dual rounding) where we lose only a factor η in the objective value as compared to the optimal (fractional) dual solution. We then use this approximate integral dual to construct a scenario $\mathcal{W}$ with cost greater than γ that gives us the contradiction. Details of the proof are provided below.

For all $k \in \mathcal{J}$, recall $\boldsymbol{v}_k$ the optimal solution corresponding to $z(\boldsymbol{e}_k)$. We have $\|\boldsymbol{v}_k\|_0 = 1$, i.e. $z(\boldsymbol{e}_k) = d_\ell v_{k\ell}$ where

$$\ell = \underset{\substack{j=1,\ldots,n \\ B_{kj} \neq 0}}{\arg\min} \frac{d_j}{B_{kj}}.$$

In particular, we have for all $j \in [n]$ such that $B_{kj} \neq 0$,

$$\frac{d_j}{B_{kj}} \geq \frac{d_\ell}{B_{k\ell}} = d_\ell v_{k\ell} = z(\boldsymbol{e}_k) > \eta\gamma w_k,$$

i.e., for all $j \in [n]$,

$$d_j \geq \eta\gamma \cdot \max_{k \in \mathcal{J}} (w_k B_{kj}).$$

For $j \in [n]$, denote

$$\hat{d}_j = \frac{d_j}{\eta\gamma \cdot \max_{k \in \mathcal{J}} (w_k B_{kj})},$$

and for all $i \in \mathcal{J}, j \in [n]$,

$$\hat{B}_{ij} = \frac{w_i B_{ij}}{\max_{k \in \mathcal{J}} (w_k B_{kj})}.$$

In particular, we have for all $i \in \mathcal{J}, j \in [n]$, $\hat{B}_{ij} \in [0,1]$ and for all $j \in [n]$, $\hat{d}_j > 1$. For any

$\mathcal{W} \subseteq \mathcal{J}$, consider the following problem

$$\hat{z}(\mathcal{W}) = \min_{\boldsymbol{y}\geq\boldsymbol{0}} \left\{ \hat{\boldsymbol{d}}^T \boldsymbol{y} \;\middle|\; \hat{\boldsymbol{B}}\boldsymbol{y} \geq \sum_{i\in\mathcal{W}} w_i \boldsymbol{e}_i \right\}. \tag{3.6}$$

We show that for any $\mathcal{W} \subseteq \mathcal{J}$, $\hat{z}(\mathcal{W})$ is just a scaling of $z(\mathcal{W})$. In particular, we have,

Claim 3.2.6. $z(\mathcal{W}) = \eta\gamma \cdot \hat{z}(\mathcal{W})$.

We present the proof of Claim 3.2.6 in Appendix C.1. To show that $z(\mathcal{J}) \leq \eta\gamma$, we show equivalently that $\hat{z}(\mathcal{J}) \leq 1$. For the sake of contradiction, suppose that $\hat{z}(\mathcal{J}) > 1$. Our goal is to construct a scenario that contradicts condition 2 of the lemma. We use ideas on dual rounding and randomized solutions from [24] and [21]. In particular, let the dual problem of $\hat{z}(\mathcal{J})$ be

$$\hat{\Delta}_{\mathcal{J}} = \max_{\boldsymbol{z}\geq\boldsymbol{0}} \left\{ \sum_{i\in\mathcal{J}} w_i z_i \;\middle|\; \sum_{i\in\mathcal{J}} \hat{B}_{ij} z_i \leq \hat{d}_j \quad \forall j \in [n] \right\}. \tag{3.7}$$

Denote $\boldsymbol{z}^*$ the optimal solution for the dual problem (3.7). By strong LP duality, we have

$$\hat{\Delta}_{\mathcal{J}} = \hat{z}(\mathcal{J}),$$

and therefore,

$$\hat{\Delta}_{\mathcal{J}} = \sum_{i\in\mathcal{J}} w_i z_i^* > 1.$$

We define the following randomized solution for all $i \in \mathcal{J}$,

$$Z_i = \lfloor z_i^* \rfloor + \xi_i,$$

where ξ_i, for $i \in \mathcal{J}$ are independent Bernoulli variables with parameter $z_i^* - \lfloor z_i^* \rfloor$, i.e.,

$$\xi_i = \mathsf{Ber}(z_i^* - \lfloor z_i^* \rfloor).$$

Claim 3.2.7. *With probability at least* $1 - O(1/n)$,

$$\left(\frac{2Z_i}{\eta}, i \in \mathcal{J}\right),$$

is a feasible solution to the dual problem (3.7).

We show that $\left(\frac{2Z_i}{\eta}, i \in \mathcal{J}\right)$ satisfies the constraints of (3.7) with high probability by using Chernoff bound concentration inequalities. The proof of Claim 3.2.7 is presented in Appendix C.1. Furthermore, we show that the cost of our randomized solution $\left(\frac{2Z_i}{\eta}, i \in \mathcal{J}\right)$ is greater than $\frac{1}{\eta}$ with a constant probability. In particular, we have the following claim.

Claim 3.2.8. $\mathbb{P}\left(\sum_{i\in\mathcal{J}} w_i Z_i > \frac{1}{2}\right) \geq 1 - e^{-\frac{1}{8}}$.

We use a concentration bound to prove Claim 3.2.8. The proof is presented in Appendix C.1. Putting Claim 3.2.7 and Claim 3.2.8 together, we have that

$$\left(\frac{2Z_i}{\eta}, i \in \mathcal{J}\right),$$

is feasible for (3.7) with high probability and has a cost $\sum_{i\in\mathcal{J}} w_i \frac{2Z_i}{\eta}$ strictly greater than $\frac{1}{\eta}$ with a non-zero constant probability. Therefore, there exists a deterministic solution for problem (3.7) with a cost at least $\frac{1}{\eta}$. For simplicity of notations, let us assume that $\left(\frac{2Z_i}{\eta}, i \in \mathcal{J}\right)$ is such a solution. Let us order $w_i Z_i$ in an increasing order, i.e.,

$$w_{(1)}Z_{(1)} \geq w_{(2)}Z_{(2)} \geq \ldots \geq w_{(|\mathcal{J}|)}Z_{(|\mathcal{J}|)}.$$

We know that $\sum_{i\in\mathcal{J}} w_i Z_i > \frac{1}{2}$. Denote L the index such that

$$\sum_{i=1}^{L-1} w_{(i)}Z_{(i)} \leq \frac{1}{2} \qquad \text{and} \qquad \sum_{i=1}^{L} w_{(i)}Z_{(i)} > \frac{1}{2}.$$

Note that $Z_{(i)}$ are integral and $Z_{(L)} \neq 0$. Hence for all $i = 1, \ldots, L$, $Z_{(i)} \geq 1$. Therefore,

$$\sum_{i=1}^{L-1} w_{(i)} \leq \sum_{i=1}^{L-1} w_{(i)} Z_{(i)} < \frac{1}{2}.$$

Note that if $L = 1$, $\sum_{i=1}^{L} w_{(i)} = w_{(1)} \leq 1$ because all w_i are in $[0, 1]$. On the other hand, if $L \geq 2$, then

$$w_{(L)} \leq w_{(L)} Z_{(L)} \leq w_{(1)} Z_{(1)} \leq \sum_{i=1}^{L-1} w_{(i)} Z_{(i)} < \frac{1}{2},$$

and therefore,

$$\sum_{i=1}^{L} w_{(i)} = \sum_{i=1}^{L-1} w_{(i)} + w_{(L)} \leq \frac{1}{2} + \frac{1}{2} = 1.$$

Therefore, in both cases we have $\sum_{i=1}^{L} w_{(i)} \leq 1$. Denote $\mathcal{W} \subseteq \mathcal{J}$ the set of indices corresponding to the top L of $w_{(i)} Z_{(i)}$. In particular, we have,

$$\sum_{i \in \mathcal{W}} w_i \leq 1.$$

Note that $\sum_{i \in \mathcal{W}} w_i Z_i > \frac{1}{2}$, and consequently,

$$\sum_{i \in \mathcal{W}} w_i \frac{2Z_i}{\eta} > \frac{1}{\eta}.$$

Consider the following problem

$$\hat{\Delta}_{\mathcal{W}} = \max_{\boldsymbol{z} \geq \boldsymbol{0}} \left\{ \sum_{i \in \mathcal{W}} w_i z_i \;\middle|\; \sum_{i \in \mathcal{W}} \hat{B}_{ij} z_i \leq \hat{d}_j \quad \forall j \in [n] \right\} \tag{3.8}$$

and its dual,

$$\hat{z}(\mathcal{W}) = \min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \hat{\boldsymbol{d}}^T \boldsymbol{y} \;\middle|\; \hat{\boldsymbol{B}} \boldsymbol{y} \geq \sum_{i \in \mathcal{W}} w_i \boldsymbol{e}_i \right\}. \tag{3.9}$$

We have shown the existence of a solution $(\frac{2Z_i}{\eta})_{i \in \mathcal{W}}$ to problem (3.8) with a cost strictly greater than $\frac{1}{\eta}$. Hence, by LP duality

$$\hat{z}(\mathcal{W}) > \frac{1}{\eta}.$$

Note that $z(\mathcal{W}) = \eta\gamma \cdot \hat{z}(\mathcal{W})$. Hence, $z(\mathcal{W}) > \gamma$ which contradicts condition 2 of our lemma.

3.3 Intersection of disjoint budget constraints.

In this section, we consider more general uncertainty sets that are defined by intersection of budget constraints that model many practical settings. We first consider the case where the budget constraints are *disjoint*. In particular, consider $S_1, S_2, \ldots, S_L$ a partition of $\{1, 2, \ldots, m\}$, i.e.

$$\bigcup_{\ell=1}^{L} S_\ell = [m] \qquad \text{and} \qquad S_i \cap S_j = \emptyset, \ \forall i \neq j.$$

We define *disjoint constrained budgeted sets* as follows

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \ \Bigg| \ \sum_{i \in S_\ell} w_{\ell i} h_i \leq 1 \ \ \forall \ell \in [L] \right\}. \tag{3.10}$$

where $S_\ell, \ell = 1, \ldots, L$ is a partition of $[m]$. This is an important class of uncertainty sets that generalizes the budget of uncertainty set (1.5). These are essentially Cartesian product of L budget of uncertainty sets. A special case of this class of uncertainty sets where all $w_{\ell i}$ are equal has been considered for example in Gupta et al. [51] and Feige et al. [24]. Recall for $L = 1$, namely the budget of uncertainty set (1.5), affine policy gives the optimal approximation to (1.1) (see Theorem 3.2.1). Our result in this section show that the performance of affine policy remains near-optimal for the more general class (3.10). In particular, we have the following theorem.

Theorem 3.3.1. *Consider the two-stage adjustable problem* (1.1) *where* $\mathcal{U}$ *is the disjoint constrained budgeted set* (3.10). *Then,*

$$z_{\mathsf{Aff}}(\mathcal{U}) = O\left(\frac{\log^2 n}{\log\log n}\right) \cdot z_{\mathsf{AR}}(\mathcal{U}).$$

Our analysis relies on constructing a feasible affine solution for (1.1) and relating the worst-case performance to a lower bound of (1.1). In particular, we consider the *online fractional* *covering problem* and use an online algorithm with $O(\log n)$- *competitive ratio* to both construct a feasible affine and also a lower bound for (1.1). The performance bound of our feasible affine is related to the competitive ratio of the online algorithm. We first introduce some preliminaries before discussing our construction and analysis.

3.3.1 Online fractional covering.

Recall, for $i = 1, \ldots, m$, $\alpha_i = 1 - (\boldsymbol{A}\boldsymbol{x}^*)_i$. We consider $\tilde{\boldsymbol{B}} \in \mathbb{R}_+^{m \times n}$ where the i-th row $\tilde{\boldsymbol{B}}_i^T = \boldsymbol{B}_i^T / \alpha_i$ if $\alpha_i > 0$ and $\tilde{\boldsymbol{B}}_i^T = \boldsymbol{B}_i^T$ otherwise. Note that $\tilde{\boldsymbol{B}}$ is a non-negative matrix. We consider the (offline) fractional covering problem

$$\Theta(\boldsymbol{h}) = \min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \mid \tilde{\boldsymbol{B}} \boldsymbol{y} \geq \boldsymbol{h} \right\},$$

for any requirement $\boldsymbol{h} \in \{0, 1\}^m$. The *online fractional covering problem* is an online version of the covering problem where the requirements are revealed online in a sequential manner. In particular, at each step we get a new constraint $\sum_{j=1}^n \tilde{B}_{ij} y_j \geq 1$ for some i and the algorithm needs to augment the current solution to satisfy the new requirement in each step.

This problem has been studied in the literature. We refer the reader to Buchbinder and Naor [52] for an extensive discussion of the problem. Buchbinder and Naor [53] give an online algorithm $\mathcal{A}$ for the online fractional covering problem that is $O(\log n)$-*competitive* (see Theorem 4.1 in [53]). In other words, the cost of the solution given by $\mathcal{A}$ for any set and sequence of requirements is at most $O(\log n)$ times the cost of the optimal solution of the corresponding offline covering problem where all the requirements are known upfront. In particular, for any sequence of requirements τ, we have

$$\max_{\tau} \frac{\mathcal{A}(\tau)}{\Theta(\tau)} = O(\log n),$$

where $\mathcal{A}(\tau)$ is the online covering cost and $\Theta(\tau)$ is the offline covering cost. Note that the competitive ratio guarantee also holds for the case where in each step, we get a subset of constraints instead of just a single constraint. We consider the algorithm $\mathcal{A}$ of Buchbinder and Naor [53] for our analysis. Let us introduce some notations that we will use for our construction and analysis.

Notations. Consider a sequence of subsets of constraints given by $(S_1, \ldots, S_R)$ where $S_r \subseteq [m]$ and $S_r \cap S_{r'} = \emptyset$ for all $r \neq r'$. In particular, in step r we get subset S_r of constraints

$$\sum_{j=1}^{n} \tilde{B}_{ij} y_j \geq 1 \qquad \forall i \in S_r.$$

For brevity of notations, let

$$\boldsymbol{h}_r = \mathbb{1}(S_r).$$

In particular, the sequence $(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_R)$ verifies $\boldsymbol{h}_r \in \{0,1\}^m$ for all $r \in [R]$ and $\sum_{r=1}^{R} \boldsymbol{h}_r \leq \boldsymbol{e}$. We introduce the following definitions.

1. **Online cost.** We denote $\mathcal{A}(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r)$ the (online) cost of covering the sequence $(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r)$ using the online algorithm $\mathcal{A}$.

2. **Online augmenting cost.** We denote $\mathcal{A}(\boldsymbol{h}_{r+1} | \ \boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r)$ the extra cost to cover $\boldsymbol{h}_{r+1}$ using the online algorithm $\mathcal{A}$ when the algorithm have already covered the sequence $(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r)$. By definition, the online augmenting cost is given by

$$\mathcal{A}(\boldsymbol{h}_{r+1} | \boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r) = \mathcal{A}(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r, \boldsymbol{h}_{r+1}) - \mathcal{A}(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r). \tag{3.11}$$

3. **Greedy augmenting cost.** We denote $\mathsf{Aug}(\boldsymbol{h}_{r+1} | \ \boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r)$ the optimal cost to cover $\boldsymbol{h}_{r+1}$ given that the sequence $(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r)$ was already covered by the online algorithm $\mathcal{A}$. In particular, the greedy augmenting cost is given by

$$\mathsf{Aug}(\boldsymbol{h}_{r+1} | \ \boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r) = \min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \mid \tilde{\boldsymbol{B}} \left(\boldsymbol{y} + \boldsymbol{y}_r^{\mathcal{A}} \right) \geq \boldsymbol{h}_{r+1} \right\}, \tag{3.12}$$

   where $\boldsymbol{y}_r^{\mathcal{A}}$ is the online solution corresponding to the cost $\mathcal{A}(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r)$.

4. **Offline cost.** Denote $\Theta(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r)$ the optimal (offline) cost to cover $(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r)$ i.e.,

$$\Theta(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r) = \Theta\left(\sum_{i=1}^{r} \boldsymbol{h}_i\right) = \min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \tilde{\boldsymbol{B}} \boldsymbol{y} \geq \sum_{i=1}^{r} \boldsymbol{h}_i \right\}.$$

Since $\mathcal{A}$ is $O(\log n)$-competitive. Then for any sequence $(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r)$,

$$\mathcal{A}(\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_r) = O(\log n) \cdot \Theta\left(\sum_{i=1}^{r} \boldsymbol{h}_i\right). \tag{3.13}$$

3.3.2 Construction of our affine solution.

Similar to the proof of Theorem 3.2.1, we construct a feasible affine solution where we split the components of $[m]$ into two subsets and cover one using a linear solution and the remaining components using a static solution. Consider $1, 2, \ldots, L$ the blocks of components of the the disjoint constrained budgeted set (3.10). For each block of components, we construct a threshold using the online fractional covering algorithm $\mathcal{A}$. This threshold defines the *expensive* components that we cover using a static solution and *inexpensive* components that we cover using a linear solution.

Construction of thresholds. Let

$$\mathcal{T} = \left\{ i \in [m] \;\middle|\; \alpha_i \leq 0 \right\},$$
$$\mathcal{T}^c = [m] \setminus \mathcal{T}.$$

For $\ell = 1, \ldots, L$, denote

$$\hat{S}_\ell = S_\ell \cap \mathcal{T}^c.$$

Let us define the following sets for all $\ell \in [L]$,

$$\mathcal{U}_\ell = \left\{ \boldsymbol{h} \in \{0,1\}^m \;\middle|\; \sum_{i \in \hat{S}_\ell} w_{\ell i} h_i \leq 1 \;\text{ and }\; h_i = 0 \;\; \forall i \notin \hat{S}_\ell \right\}.$$

Note that $\oplus_{\ell=1}^{L} \mathcal{U}_\ell \subseteq \mathcal{U}$. We construct a greedy sequence $(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_L)$ where each $\boldsymbol{a}_\ell$ is chosen from some set $\mathcal{U}_\ell$ such that it maximizes the online augmenting cost (3.11) of the sequence. Algorithm 1 describes the procedure in details.

Algorithm 1 Computing a greedy scenario $\boldsymbol{a}$

1: Initialize $\mathcal{L} = \{1, 2, \ldots, L\}$.
2: **for** $\ell = 1, \ldots, L$ **do**
3: $\quad (s, \boldsymbol{b}) = \underset{s \in \mathcal{L},\ \boldsymbol{b} \in \mathcal{U}_s}{\arg\max}\ \mathcal{A}(\boldsymbol{b} \mid \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{\ell-1})$
4: $\quad$ Set $\boldsymbol{a}_\ell = \boldsymbol{b}$, update $\mathcal{L} = \mathcal{L} \setminus s$
5: **end for**

Algorithm 1 constructs the greedy sequence when the covering constraint matrix is $\tilde{\boldsymbol{B}}$ and therefore, the guarantee of the online algorithm $\mathcal{A}$ gives us a bound of $O(\log n)$ between the online cost, $\mathcal{A}(\boldsymbol{a})$ and offline cost, $\Theta(\boldsymbol{a})$ of the covering problem with $\tilde{\boldsymbol{B}}$ for the greedy sequence, $\boldsymbol{a}$. The following lemma relates this cost with OPT for (1.1).

Lemma 3.3.2. *For all $\ell \in [L]$, denote v_ℓ the cost of covering the sequence $(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_\ell)$ using the online algorithm $\mathcal{A}$, i.e.,*

$$v_\ell = \mathcal{A}(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_\ell).$$

We have

$$v_L = O(\log n) \cdot \text{OPT}.$$

Proof. We suppose wlog that the sequence $(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_L)$ belongs respectively to $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_L$ and let

$$\boldsymbol{a} = \sum_{\ell=1}^{L} \boldsymbol{a}_\ell \in \mathcal{U}.$$

Then,

$$\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{a}) \geq \boldsymbol{a}.$$

Denote $\boldsymbol{a} = \sum_{i \in \mathcal{S}} \boldsymbol{e}_i$. Hence,

$$\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{a}) \geq \sum_{i \in \mathcal{S}} \boldsymbol{e}_i - \sum_{i=1}^{m} (1 - \alpha_i)\boldsymbol{e}_i,$$

i.e.,

$$\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{a}) \geq \sum_{i\in\mathcal{S}} \alpha_i \boldsymbol{e}_i - \sum_{i\notin\mathcal{S}} (1-\alpha_i)\boldsymbol{e}_i.$$

Moreover, $\boldsymbol{B}$ and $\boldsymbol{y}^*(\boldsymbol{a})$ are non-negative. Hence, $\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{a}) \geq \boldsymbol{0}$ and therefore,

$$\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{a}) \geq \sum_{i\in\mathcal{S}} \alpha_i \boldsymbol{e}_i,$$

which is equivalent to

$$\tilde{\boldsymbol{B}}\boldsymbol{y}^*(\boldsymbol{a}) \geq \sum_{i\in\mathcal{S}} \boldsymbol{e}_i = \boldsymbol{a}.$$

Therefore,

$$\Theta(\boldsymbol{a}) = \min_{\boldsymbol{y}\geq\boldsymbol{0}} \left\{ \boldsymbol{d}^T\boldsymbol{y} \;\middle|\; \tilde{\boldsymbol{B}}\boldsymbol{y} \geq \boldsymbol{a} \right\} \leq \boldsymbol{d}^T y^*(\boldsymbol{a}) \leq \text{OPT}.$$

Finally,

$$\nu_L = O(\log n)\cdot\Theta(\boldsymbol{a}) = O(\log n)\cdot \text{OPT},$$

where the first equality follows from (3.13). □

Now, we are ready to construct our feasible affine solution that has a cost $O\left(\frac{\log^2 n}{\log\log n}\right)$ times $z_{\mathsf{AR}}(\mathcal{U})$ using a linear and a static part. Recall for all $i \in [m]$, $z(\boldsymbol{e}_i)$ the cost of covering component $\boldsymbol{e}_i$ in the second stage as defined in (3.2) and $\boldsymbol{v}_i$ an optimal corresponding solution. For all $\ell = 1,\ldots,L$, we consider the following sets of components

$$\mathcal{I}_\ell = \left\{ i \in S_\ell \;\middle|\; \alpha_i > 0 \;\text{ and }\; \frac{\alpha_i z(\boldsymbol{e}_i)}{w_{\ell i}} \leq \beta\cdot(\nu_\ell - \nu_{\ell-1}) \right\},$$

where

$$\beta = \frac{8\log n}{\log\log n},$$

and

$$\nu_\ell - \nu_{\ell-1} = \mathcal{A}(\boldsymbol{a}_\ell|\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_{\ell-1}),$$

as defined in (3.11). Denote

$$\mathcal{I} = \bigcup_{\ell=1}^{L} \mathcal{I}_\ell.$$

and $\mathcal{I}^c$ its complement, i.e., $\mathcal{I}^c = [m] \setminus \mathcal{I}$.

Linear part. We cover a fraction of the components of $\mathcal{I}$ using the following linear solution for any $\boldsymbol{h} \in \mathcal{U}$,

$$\boldsymbol{y}_{\mathsf{Lin}}(\boldsymbol{h}) = \sum_{i \in \mathcal{I}} \alpha_i h_i \boldsymbol{v}_i. \tag{3.14}$$

Static part. We use a static solution to cover the remaining components $\boldsymbol{e}_i$ where $i \in \mathcal{I}^c$ and $(1 - \alpha_i)^+ \boldsymbol{e}_i$ for $i \in \mathcal{I}$. In particular, similar to (3.4) we consider the following static problem

$$(\boldsymbol{x}_{\mathsf{Sta}}, \boldsymbol{y}_{\mathsf{Sta}}) \in \arg\min_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \sum_{i \in \mathcal{I}^c} \boldsymbol{e}_i + \sum_{i \in \mathcal{I}} (1 - \alpha_i)^+ e_i \right\}, \tag{3.15}$$

and denote $z_{\mathsf{Sta}} = \boldsymbol{c}^T \boldsymbol{x}_{\mathsf{Sta}} + \boldsymbol{d}^T \boldsymbol{y}_{\mathsf{Sta}}$. Our affine solution is given by

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{x}_{\mathsf{Sta}} \\ \boldsymbol{y}(\boldsymbol{h}) &= \boldsymbol{y}_{\mathsf{Lin}}(\boldsymbol{h}) + \boldsymbol{y}_{\mathsf{Sta}}, \qquad \forall \boldsymbol{h} \in \mathcal{U}. \end{aligned} \tag{3.16}$$

We can show that the affine solution (3.16) is feasible for (1.1). In particular, we have

Lemma 3.3.3 (Feasibility). *The affine solution in* (3.16) *is feasible for the adjustable problem* (1.1).

The proof is similar to the proof of Lemma 3.2.2.

3.3.3 Cost analysis.

In the following two lemmas, we analyze the cost of the linear and static parts in our affine solution (3.16).

Lemma 3.3.4 (Cost of Linear part). *The cost of the linear part* $\boldsymbol{y}_{\mathsf{Lin}}(\boldsymbol{h})$ *defined in* (3.16) *is* $O(\beta \log n) \cdot \mathrm{OPT}$ *for any* $\boldsymbol{h} \in \mathcal{U}$.

Proof. We have,

$$\begin{aligned}
\boldsymbol{d}^T \boldsymbol{y}_{\mathsf{Lin}}(\boldsymbol{h}) = \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}_\ell} \alpha_i h_i \boldsymbol{d}^T \boldsymbol{v}_i = \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}_\ell} \alpha_i h_i z(\boldsymbol{e}_i) &\le \sum_{\ell=1}^{L} \beta \cdot (\nu_\ell - \nu_{\ell-1}) \sum_{i \in \mathcal{I}_\ell} w_{\ell i} h_i \\
&\le \beta \sum_{\ell=1}^{L} (\nu_\ell - \nu_{\ell-1}) \\
&= \beta \cdot \nu_L \\
&= O(\beta \log n) \cdot \mathrm{OPT}.
\end{aligned}$$

where the first inequality holds because $\alpha_i z(\boldsymbol{e}_i) \le \beta w_{\ell i} \cdot (\nu_\ell - \nu_{\ell-1})$ for all $i \in \mathcal{I}_\ell$ and $\ell \in [L]$, the second inequality holds because $\sum_{i \in \mathcal{I}_\ell} w_{\ell i} h_i \le 1$ for any $\boldsymbol{h} \in \boldsymbol{\mathcal{U}}$ and the last equality follows from Lemma 3.3.2.

□

Lemma 3.3.5 (Cost of Static part). *The cost of the static part* $(\boldsymbol{x}_{\mathsf{Sta}}, \boldsymbol{y}_{\mathsf{Sta}})$ *defined in* (3.15) *is* $O(\beta \log n) \cdot \mathrm{OPT}$.

Proof. Denote $\boldsymbol{y}_{\mathcal{A}}$ the solution provided by the online algorithm $\mathcal{A}$ that covers the sequence $(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_L)$. Consider

$$\mathcal{J}_1 = \left\{ i \in [m] \;\middle|\; (\tilde{\boldsymbol{B}} \boldsymbol{y}_{\mathcal{A}})_i \ge \frac{1}{2} \text{ and } \alpha_i > 0 \right\}.$$

We have for all $i \in \mathcal{J}_1$, $2\boldsymbol{B}\boldsymbol{y}_{\mathcal{A}} \ge \alpha_i \boldsymbol{e}_i$, i.e.,

$$2\boldsymbol{B}\boldsymbol{y}_{\mathcal{A}} \ge \sum_{i \in \mathcal{J}_1} \alpha_i \boldsymbol{e}_i.$$

and from Lemma 3.3.2,

$$2\boldsymbol{d}^T \boldsymbol{y}_{\mathcal{A}} = 2\nu_L = O(\log n) \cdot \mathrm{OPT}.$$

Now, we focus on the set of the remaining components. Denote

$$\mathcal{J}_2 = \mathcal{I}^c \setminus \{\mathcal{T} \cup \mathcal{J}_1\}$$

and for $\ell = 1, \ldots, L$ denote

$$\mathcal{V}_\ell = \mathcal{J}_2 \cap S_\ell.$$

For each $\ell \in [L]$, we apply the structural result in Lemma 3.2.4 for the subset $\mathcal{V}_\ell$ with parameters $w_{\ell i}$, $\gamma = 2(\nu_\ell - \nu_{\ell-1})$, cost vector $\boldsymbol{d}$ and constraint matrix $\tilde{\boldsymbol{B}}$. The first condition of Lemma 3.2.4 is satisfied because for any $i \in \mathcal{V}_\ell$,

$$\frac{\alpha_i z(\boldsymbol{e}_i)}{w_{\ell i}} = \frac{\Theta(\boldsymbol{e}_i)}{w_{\ell i}} > \beta(\nu_\ell - \nu_{\ell-1}) = 4\gamma \frac{\log n}{\log \log n}.$$

Consider any $\mathcal{W} \subseteq \mathcal{V}_\ell$ such that $\sum_{i \in \mathcal{W}} w_{\ell i} \leq 1$. We have $\mathbb{1}(\mathcal{W}) \in \mathcal{U}_\ell$. Moreover, $\boldsymbol{y}_{\mathcal{A}}$ covers less than $\frac{1}{2}\mathbb{1}(\mathcal{W})$. Therefore,

$$\frac{1}{2}\Theta\left(\mathcal{W}\right) \leq \mathsf{Aug}\left(\mathbb{1}(\mathcal{W}) \middle| \, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_L\right).$$

Furthermore,

$$\begin{aligned}
\mathsf{Aug}\left(\mathbb{1}(\mathcal{W}) \middle| \, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_L\right) &\leq \mathsf{Aug}\left(\mathbb{1}(\mathcal{W}) \middle| \, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{\ell-1}\right) \\
&\leq \mathcal{A}\left(\mathbb{1}(\mathcal{W}) \middle| \, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{\ell-1}\right) \\
&\leq \mathcal{A}\left(\boldsymbol{a}_\ell \middle| \, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{\ell-1}\right) \\
&= \nu_\ell - \nu_{\ell-1},
\end{aligned}$$

where the first inequality holds because the cost of covering $\mathbb{1}(\mathcal{W})$ given the online solution for $(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_L)$ is smaller than the cost of covering $\mathbb{1}(\mathcal{W})$ given the online solution for $(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{\ell-1})$. The second inequality holds because the cost of the online algorithm is less than the offline cost and the third one follows from Step 3 in the construction of the greedy scenario $\boldsymbol{a}$

in Algorithm 1. Hence,

$$\Theta(\mathcal{W}) \le 2(\nu_\ell - \nu_{\ell-1}) = \gamma.$$

Therefore, the second condition of Lemma 3.2.4 is also satisfied and consequently,

$$\Theta(\mathcal{V}_\ell) \le 4\gamma \frac{\log n}{\log\log n} = \beta(\nu_\ell - \nu_{\ell-1}).$$

By taking the sum over all $\ell = 1, \ldots, L$, we get

$$\sum_{\ell=1}^{L} \Theta(\mathcal{V}_\ell) \le \beta \nu_L = O(\beta \log n) \cdot \mathrm{OPT}.$$

For $\ell \in [L]$, denote $\boldsymbol{y}_\ell$ an optimal solution corresponding to $\Theta(\mathcal{V}_\ell)$. In particular,

$$\sum_{\ell=1}^{L} \boldsymbol{B}\boldsymbol{y}_\ell \ge \sum_{i \in \mathcal{J}_2} \alpha_i \boldsymbol{e}_i.$$

By feasibility of the optimal solution, we have

$$\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{0}) \ge \boldsymbol{0},$$

i.e.,

$$\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{0}) \ge \sum_{i=1}^{m} (\alpha_i - 1)\boldsymbol{e}_i.$$

Moreover, since $\boldsymbol{B}$ and $y^*(\boldsymbol{0})$ are non-negative, we have

$$\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{0}) \ge \sum_{i=1}^{m} (\alpha_i - 1)^+ \boldsymbol{e}_i.$$

Therefore, we have the following candidate static solution for (3.15):

$$\begin{aligned}
\boldsymbol{x}_{\mathsf{Sta}} &= \boldsymbol{x}^* \\
\boldsymbol{y}_{\mathsf{Sta}} &= \boldsymbol{y}^*(\boldsymbol{0}) + 2\boldsymbol{y}_{\mathcal{A}} + \sum_{\ell=1}^{L} \boldsymbol{y}_{\ell}.
\end{aligned}$$

Putting it all together,

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{0}) + 2\boldsymbol{B}\boldsymbol{y}_{\mathcal{A}} + \sum_{\ell=1}^{L} \boldsymbol{B}\boldsymbol{y}_{\ell} &\geq \sum_{i=1}^{m}(1-\alpha_i)\boldsymbol{e}_i + \sum_{i=1}^{m}(\alpha_i - 1)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_1}\alpha_i\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_2}\alpha_i\boldsymbol{e}_i \\
&= \sum_{i=1}^{m}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_1\cup\mathcal{J}_2}\alpha_i\boldsymbol{e}_i \\
&= \sum_{i\in\mathcal{I}}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{I}^c}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_1\cup\mathcal{J}_2}\alpha_i\boldsymbol{e}_i \\
&= \sum_{i\in\mathcal{I}}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\tau}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_1\cup\mathcal{J}_2}(1-\alpha_i)^+\boldsymbol{e}_i + \alpha_i\boldsymbol{e}_i \\
&\geq \sum_{i\in\mathcal{I}}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\tau}\boldsymbol{e}_i + \sum_{i\in\mathcal{J}_1\cup\mathcal{J}_2}\boldsymbol{e}_i \\
&= \sum_{i\in\mathcal{I}}(1-\alpha_i)^+\boldsymbol{e}_i + \sum_{i\in\mathcal{I}^c}\boldsymbol{e}_i
\end{aligned}$$

where the last inequality holds because $\alpha_i \leq 0$ for all $i \in \tau$ and $(1-\alpha_i)^+ + \alpha_i \geq 1$ for all $i \in \mathcal{J}_1 \cup \mathcal{J}_2$.

Therefore,

$$\begin{aligned}
z_{\mathsf{Sta}} &\leq \boldsymbol{c}^T\boldsymbol{x}^* + \boldsymbol{d}^T\boldsymbol{y}^*(\boldsymbol{0}) + 2\boldsymbol{d}^T\boldsymbol{y}_{\mathcal{A}} + \sum_{\ell=1}^{L} z(\mathcal{V}_{\ell}) \\
&\leq \mathrm{OPT} + O(\log n)\cdot\mathrm{OPT} + O(\beta\log n)\cdot\mathrm{OPT} \\
&= O(\beta\log n)\cdot\mathrm{OPT}.
\end{aligned}$$

□

Proof of Theorem 3.3.1. Lemma 3.3.3 show that our affine solution (3.16) is feasible for the adjustable problem (1.1). Lemma 3.3.4 and 3.3.5 show that the cost of the feasible affine solution is less than

$$O(\beta \log n) \cdot \text{OPT} + O(\beta \log n) \cdot \text{OPT} = O\left(\frac{\log^2 n}{\log \log n}\right) \cdot \text{OPT}$$

which implies that,

$$z_{\mathsf{Aff}}(\mathcal{U}) = O\left(\frac{\log^2 n}{\log \log n}\right) \cdot z_{\mathsf{AR}}(\mathcal{U}).$$

We would like to note that Gupta et al. [51] give $O(\log n)$-approximation to (1.1) in the special case $\boldsymbol{A}, \boldsymbol{B} \in \{0,1\}^{m \times n}$, $\boldsymbol{d} = \lambda \boldsymbol{c}$ and $w_{\ell i} = w$ are all ℓ, i for some constant w. Therefore, for this special case the bound of [51] is stronger than our bound in Theorem 3.3.1. However, their algorithm does not give a functional policy approximation. Here, our focus is different, namely, to analyze the performance of affine policies that are widely used in practice and exhibit strong empirical performance. Our analysis shows that the performance of affine policies for disjoint constrained budgeted sets is near-optimal and nearly matches the hardness of the problem. Note that our bound in Theorem 3.3.1 is not necessarily tight. It is an interesting open question to study if affine policies also give an optimal approximation for this more general class of budgeted uncertainty sets.

3.4 General intersection of budgeted uncertainty sets.

In this section, we consider the general intersection of budget of uncertainty sets given by (1.6)

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i \in S_\ell} w_{\ell i} h_i \leq 1 \;\; \forall \ell \in [L] \right\},$$

where $\boldsymbol{w}_\ell \in [0,1]^m$ and S_ℓ for $\ell \in [L]$ is a general family of subsets of $[m]$. This class is a generalization of the single budget of uncertainty set (1.5). It captures many important sets including CLT sets considered in Bertsimas and Bandi [37] and *inclusion-constrained budgeted* sets considered in Gounaris et al. [38]. In this section , we study the performance of affine policies for intersection of budget of uncertainty sets (1.6) and show strong theoretical guarantees. We start by the case when the set (1.6) verifies some symmetric properties (*permutation invariant sets*) and then we give our results for the general form (1.6).

3.4.1 Permutation Invariant Sets.

We consider intersection of budgeted sets that are *permutation invariant*.

Definition 3.4.1 (Permutation Invariant Sets). We say that $\mathcal{U}$ is a *permutation invariant set* if $\boldsymbol{x} \in \mathcal{U}$ implies that for any permutation τ of $\{1, 2, \ldots, m\}$, $\boldsymbol{x}^\tau \in \mathcal{U}$ where $x_i^\tau = x_{\tau(i)}$.

This class of sets captures many important sets including CLT sets that have been considred in Bertsimas and Bandi [37]. Note that a CLT set is given by

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\Big|\; \sum_{i \in \mathcal{S}} h_i \leq \Gamma \;\; \forall \mathcal{S} \subseteq [m], |\mathcal{S}| = k \right\}, \tag{3.17}$$

for some $k \in \mathbb{N}$. The following theorem gives our performance bound for affine policies under the class of intersection of budgeted sets that are permutation invariant.

Theorem 3.4.2. *Consider the two-stage adjustable problem* (1.1) *where $\mathcal{U}$ is the intersection of L budget constraints* (1.6)*. Suppose that $\mathcal{U}$ is permutation invariant set and $\mathcal{X}$ is a polyhedral cone. Then,*

$$z_{\mathsf{Aff}}(\mathcal{U}) = O\left(\log L \cdot \frac{\log n}{\log \log n}\right) \cdot z_{\mathsf{AR}}(\mathcal{U}).$$

Proof. Our proof relies on a geometric property that we show for budgeted uncertainty sets that are permutation invariant. In particular, we show that for any $\mathcal{U}$ permutation invariant, there exists a (single) budget of uncertainty set $\mathcal{V}$ of the form (1.5) such that

$$\frac{1}{4 \log L} \cdot \mathcal{V} \subseteq \mathcal{U} \subseteq 2\mathcal{V}. \tag{3.18}$$

Since $\mathcal{U}$ is permutation invariant,

$$\gamma \boldsymbol{e} \in \arg\max \left\{ \boldsymbol{e}^T \boldsymbol{h} \;\middle|\; \boldsymbol{h} \in \mathcal{U} \right\},$$

for some $\gamma \in [0, 1]$. Consider

$$\xi_i \overset{i.i.d.}{\sim} \mathsf{Ber}(\gamma) \qquad i = 1, \ldots, m,$$

i.e., $\xi_1, \xi_2, \ldots, \xi_m$ are i.i.d. Bernoulli random variables of parameter γ. Let

$$\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_m).$$

Consider the following budget of uncertainty set

$$\tilde{\mathcal{V}} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^m h_i \le \sum_{i=1}^m \xi_i \right\}.$$

Note that $\tilde{\mathcal{V}}$ is random depending on the realization of $\xi_1, \xi_2, \ldots, \xi_m$. We show that

$$\mathbb{P}\left(\frac{1}{4 \log L} \cdot \tilde{\mathcal{V}} \subseteq \mathcal{U} \subseteq 2\tilde{\mathcal{V}} \right) > \epsilon,$$

for some constant $\epsilon > 0$ which implies the existence of $\tilde{\mathcal{V}}$ such that (3.18) is verified. For that purpose, we show first that the right inclusion holds with a constant probability and then the left one holds with high probability.

Claim 3.4.3. $\mathbb{P}\left(\mathcal{U} \subseteq 2\tilde{\mathcal{V}}\right) \ge 1 - e^{-\frac{1}{8}}$.

Let us prove the above claim. Note that,

$$\gamma m = \max_{\boldsymbol{h} \in \mathcal{U}} \boldsymbol{e}^T \boldsymbol{h}.$$

Suppose that,

$$\gamma m \leq 2\boldsymbol{e}^T\boldsymbol{\xi}.$$

Then for all $\boldsymbol{h} \in \mathcal{U}$,

$$\boldsymbol{e}^T\boldsymbol{h} \leq 2\boldsymbol{e}^T\boldsymbol{\xi}$$

i.e., for all $\boldsymbol{h} \in \mathcal{U}$

$$\boldsymbol{h} \in 2\tilde{\mathcal{V}}.$$

Hence, $\gamma m \leq 2\boldsymbol{e}^T\boldsymbol{\xi}$ implies that $\mathcal{U} \subseteq 2\tilde{\mathcal{V}}$. Therefore,

$$\begin{aligned}\mathbb{P}\left(\mathcal{U} \subseteq 2\tilde{\mathcal{V}}\right) &\geq \mathbb{P}\left(2\boldsymbol{e}^T\boldsymbol{\xi} \geq \gamma m\right)\\ &= \mathbb{P}\left(\sum_{i=1}^{m}\xi_i \geq \frac{1}{2}\gamma m\right).\end{aligned}$$

We know that $\mathbb{E}\left(\sum_{i=1}^{m}\xi_i\right) = \gamma m$. Therefore, from the Chernoff inequality in Lemma C.2.2, we have,

$$\mathbb{P}\left(\sum_{i=1}^{m}\xi_i \geq \frac{1}{2}\gamma m\right) \geq 1 - \exp\left(-\frac{\gamma m}{8}\right) \geq 1 - e^{-\frac{1}{8}}.$$

where the last inequality holds because $\gamma m \geq 1$ since $\boldsymbol{e}_i \in \mathcal{U}$ for all i and $\mathcal{U}$ is convex.

Claim 3.4.4. $\mathbb{P}\left(\tilde{\mathcal{V}} \subseteq 4\log L \cdot \mathcal{U}\right) \geq 1 - \frac{1}{L}$.

Note that $\boldsymbol{\xi}$ is an extreme point of $\tilde{\mathcal{V}}$ and that all pareto extreme points of $\tilde{\mathcal{V}}$ are just permutation of $\boldsymbol{\xi}$. Moreover, we know that $\mathcal{U}$ is permutation invariant set, hence if $\mathcal{U}$ contains $\boldsymbol{\xi}$ then $\mathcal{U}$ contains all pareto extreme points of $\mathcal{U}$ and consequently contains $\mathcal{U}$ by down-monotonicity. Therefore,

$$\begin{aligned}\mathbb{P}\left(\tilde{\mathcal{V}} \subseteq 4\log L \cdot \mathcal{U}\right) &\geq \mathbb{P}\left(\boldsymbol{\xi} \in 4\log L \cdot \mathcal{U}\right)\\ &= \mathbb{P}\left(\boldsymbol{v}_\ell^T\boldsymbol{\xi} \leq 4\log L, \;\; \forall \ell \in [L]\right)\\ &= 1 - \mathbb{P}\left(\exists \ell \in [L], \; \boldsymbol{v}_\ell^T\boldsymbol{\xi} > 4\log L\right)\\ &\geq 1 - \sum_{\ell=1}^{L}\mathbb{P}\left(\boldsymbol{v}_\ell^T\boldsymbol{\xi} > 4\log L\right),\end{aligned}$$

where the last inequality follows from a union bound. We have $\mathbb{E}(\boldsymbol{v}_\ell^T \boldsymbol{\xi}) = \boldsymbol{v}_\ell^T \gamma \boldsymbol{e} \leq 1 \leq \log L$ for $L \geq 2$ because $\gamma \boldsymbol{e}$ is a feasible point in $\mathcal{U}$. Therefore, from Lemma C.2.1 with $\delta = 3$.

$$\mathbb{P}\left(\boldsymbol{v}_\ell^T \boldsymbol{\xi} > 4 \log L\right) \leq \left(\frac{e^3}{4^4}\right)^{\log L} \leq \left(e^{-2}\right)^{\log L} = \frac{1}{L^2}.$$

We conclude that,

$$\mathbb{P}\left(\tilde{\mathcal{V}} \subseteq 4 \log L \cdot \mathcal{U}\right) \geq 1 - \sum_{\ell=1}^{L} \frac{1}{L^2} = 1 - \frac{1}{L}.$$

Hence, from Claim 3.4.3 and Claim 3.4.4, there exists a budget of uncertainty set $\tilde{\mathcal{V}}$ with a non zero probability that verifies the inclusion in (3.18). Therefore,

$$\begin{aligned} z_{\mathsf{Aff}}(\mathcal{U}) &\leq 2 \cdot z_{\mathsf{Aff}}(\tilde{\mathcal{V}}) \\ &= 2 \cdot O\left(\frac{\log n}{\log \log n}\right) \cdot z_{\mathsf{AR}}(\tilde{\mathcal{V}}) \\ &\leq 2 \cdot O\left(\frac{\log n}{\log \log n}\right) \cdot 4 \log L \cdot z_{\mathsf{AR}}(\mathcal{U}) = O\left(\frac{\log n}{\log \log n} \cdot \log L\right) \cdot z_{\mathsf{AR}}(\mathcal{U}), \end{aligned}$$

where the first inequality holds because $\mathcal{U} \subseteq 2\tilde{\mathcal{V}}$ and $2 \cdot \mathcal{X} \subseteq \mathcal{X}$ ($\mathcal{X}$ is a polyhedral cone). The first equality follows from Theorem 3.2.1 because $\tilde{\mathcal{V}}$ is a budget of uncertainty set, and finally the last inequality holds because $\tilde{\mathcal{V}} \subseteq 4 \log L \cdot \mathcal{U}$ and $4 \log L \cdot \mathcal{X} \subseteq \mathcal{X}$ ($\mathcal{X}$ is a polyhedral cone). □

We would like to note that the result of Theorem 3.4.2 extends as well to the class of intersection of budgeted sets that are *scaled permutation invariant*. We say that $\mathcal{U}$ is a scaled permutation invariant set if there exists $\boldsymbol{\lambda} \in \mathbb{R}_+^m$ and $\mathcal{V}$ a permutation invariant set such that

$$\mathcal{U} = \mathsf{diag}(\boldsymbol{\lambda}) \cdot \mathcal{V}.$$

In fact, for a given scaled permutation invariant set $\mathcal{U}$, it is possible to scale the two-stage adjustable problem (1.1) and get a new problem where the uncertainty set is permutation invariant. Indeed, suppose $\mathcal{U} = \mathsf{diag}(\boldsymbol{\lambda}) \cdot \mathcal{V}$ where $\mathcal{V}$ is a permutation invariant set; by multiplying the constraint matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ by $\mathsf{diag}(\boldsymbol{\lambda})^{-1}$, we get a new problem where the uncertainty set now is permutation invariant. The performance of affine policy is not affected by this scaling and the bound given by Theorem 3.4.2 still hold.

3.4.2 General intersection of budgets.

Consider the general intersection of budgeted sets (1.6) which is given by

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i \in S_\ell} w_{\ell i} h_i \leq 1 \;\; \forall \ell \in [L] \right\}.$$

We show that affine policy gives a worst-case bound of $O\left(L \cdot \frac{\log n}{\log \log n}\right)$ where L is the number of constraints in $\mathcal{U}$. In particular, we have the following theorem.

Theorem 3.4.5. *Consider the two-stage adjustable problem* (1.1) *where* $\mathcal{U}$ *is the intersection of L budgeted sets given by* (1.6)*. Suppose that* $\mathcal{X}$ *is a polyhedral cone. Then,*

$$z_{\mathsf{Aff}}(\mathcal{U}) = O\left(L \cdot \frac{\log n}{\log \log n}\right) \cdot z_{\mathsf{AR}}(\mathcal{U}).$$

Proof. Denote for all $\ell \in [L]$,

$$\boldsymbol{w}_\ell = \sum_{i \in S_\ell} w_{\ell i} \boldsymbol{e}_i$$

and let

$$\bar{\boldsymbol{w}} = \frac{1}{L} \sum_{\ell=1}^{L} \boldsymbol{w}_\ell.$$

Consider the following budget of uncertainty set,

$$\mathcal{V} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \bar{\boldsymbol{w}}^T \boldsymbol{h} \leq 1 \right\}.$$

We show that $\mathcal{U} \subseteq \mathcal{V} \subseteq L \cdot \mathcal{U}$. Suppose $\boldsymbol{h} \in \mathcal{U}$. Then, for any $\ell \in [L]$, we have $\boldsymbol{w}_\ell^T \boldsymbol{h} \leq 1$.

Therefore, by summing up all these inequalities and dividing by L, we get $\bar{\boldsymbol{w}}^T \boldsymbol{h} \leq 1$, i.e., $\boldsymbol{h} \in \mathcal{V}$. Hence $\mathcal{U} \subseteq \mathcal{V}$. Conversely, suppose $\boldsymbol{h} \in \mathcal{V}$. For any $\ell \in [L]$, we have $\boldsymbol{w}_\ell^T \boldsymbol{h} \leq \sum_{\ell=1}^{L} \boldsymbol{w}_\ell^T \boldsymbol{h} \leq L$, hence $\boldsymbol{h} \in L \cdot \mathcal{U}$ and consequently $\mathcal{V} \subseteq L \cdot \mathcal{U}$. Therefore,

$$
\begin{aligned}
z_{\mathsf{Aff}}(\mathcal{U}) &\leq z_{\mathsf{Aff}}(\mathcal{V}) \\
&= O\left(\frac{\log n}{\log \log n}\right) \cdot z_{\mathsf{AR}}(\mathcal{V}) \\
&\leq O\left(\frac{\log n}{\log \log n}\right) \cdot L \cdot z_{\mathsf{AR}}(\mathcal{U}),
\end{aligned}
$$

where the first inequality holds because $\mathcal{U} \subseteq \mathcal{V}$, the second one is a consequence of Theorem 3.2.1 since $\mathcal{V}$ is a budget of uncertainty set of the form (1.5), and finally the last inequality holds because $\mathcal{V} \subseteq L \cdot \mathcal{U}$ and $L \cdot \mathcal{X} \subseteq \mathcal{X}$ ($\mathcal{X}$ is a polyhedral cone).

□

3.5 Faster algorithm for near-optimal affine solutions.

In this section, we present an algorithm to compute an approximate affine policy for (1.1) under budget of uncertainty sets, that is significantly faster than solving the optimization program (A.1) that computes the optimal affine policy. Our algorithm is based on the analysis of the performance of affine policies that shows the existence of a good affine solution that satisfies certain nice structural properties. In particular, our construction of approximate affine solution in Section 3.1 partitions the components into *expensive* and *inexpensive* components based on a threshold. We cover a fraction of the *inexpensive* components using a linear solution and the remaining components using a static solution. In particular, we show that there exists an affine solution with such a structure and cost at most $O(\log n / \log \log n)$ times the optimal optimal cost of (1.1) for some partition of components into *expensive* and *inexpensive*. Based on this structure, we give a faster algorithm to compute an approximate affine solution for budget of uncertainty sets.

3.5.1 Our algorithm.

Let $\mathcal{U}$ be the budget of uncertainty set (1.5) given by

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} w_i h_i \leq 1 \right\}.$$

Recall from the proof of Theorem 3.2.1, we construct our candidate affine solution by partitioning the components of $[m]$ into two subsets $\mathcal{I}$ and its complement $\mathcal{I}^c$. The linear solution (3.3) is given by

$$\boldsymbol{y}_{\mathsf{Lin}}(\boldsymbol{h}) = \sum_{i \in \mathcal{I}} \alpha_i h_i \boldsymbol{v}_i$$

where

$$\mathcal{I} = \left\{ i \in [m] \;\middle|\; \frac{\alpha_i z(\boldsymbol{e}_i)}{w_i} \leq \beta \cdot \mathrm{OPT} \right\},$$

and for all $i \in [m]$,

$$\alpha_i = 1 - (\boldsymbol{A}\boldsymbol{x}^*)_i,$$
$$\boldsymbol{v}_i \in \arg\min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{e}_i \right\}.$$

Let

$$\boldsymbol{Y} = [\boldsymbol{v}_1 \mid \boldsymbol{v}_2 \mid \ldots \mid \boldsymbol{v}_m].$$

Based on the structure of the linear part, we propose the following approximate affine solution:

$$\boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{Y} \cdot \mathsf{diag}(\boldsymbol{\alpha}) \cdot \boldsymbol{h} + \boldsymbol{q}$$

where $\boldsymbol{Y}$ is a constant that can be computed efficiently upfront and α_i for $i \in [m]$ are non-negative variables. This structure captures our candidate solution (3.3). Hence, we reduce the number of second stage variables from $O(nm)$ in (1.2) to $O(n+m)$. Moreover, the non-negativity constraint on $\boldsymbol{y}(\boldsymbol{h})$ reduces to $\boldsymbol{\alpha} \geq \boldsymbol{0}$ and $\boldsymbol{q} \geq \boldsymbol{0}$ in this special class of affine solutions. Restricting to the above class of affine solutions, we have the following optimization problem.

$$
\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{q}} \ & \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \ \boldsymbol{d}^T \left(\boldsymbol{Y}\cdot \mathsf{diag}(\boldsymbol{\alpha})\cdot \boldsymbol{h} + \boldsymbol{q}\right) \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\left(\boldsymbol{Y}\cdot \mathsf{diag}(\boldsymbol{\alpha})\cdot \boldsymbol{h} + \boldsymbol{q}\right) \ \geq \ \boldsymbol{h}, \quad \forall \boldsymbol{h}\in\mathcal{U} \\
& \boldsymbol{x}\in\mathcal{X}, \boldsymbol{\alpha}\in\mathbb{R}^m_+, \boldsymbol{q}\in\mathbb{R}^n_+.
\end{aligned}
\tag{3.19}
$$

Using similar reformulations as in Lemma (A.0.2), the above problem can be formulated as the following LP:

$$
\begin{aligned}
\min \ & \boldsymbol{c}^T\boldsymbol{x} + z \\
& z - \boldsymbol{d}^T\boldsymbol{q} \geq \boldsymbol{r}^T\boldsymbol{v} \\
& \boldsymbol{R}^T\boldsymbol{v} \geq \boldsymbol{Y}\cdot \mathsf{diag}(\boldsymbol{\alpha})^T\boldsymbol{d} \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{q} \geq \boldsymbol{V}^T\boldsymbol{r} \\
& \boldsymbol{R}^T\boldsymbol{V} \geq \boldsymbol{I}_m - \boldsymbol{B}\boldsymbol{Y}\cdot \mathsf{diag}(\boldsymbol{\alpha}) \\
& \boldsymbol{x}\in\mathcal{X}, \ \boldsymbol{v}\in\mathbb{R}^L_+, \ \boldsymbol{U}\in\mathbb{R}^{L\times n}_+, \ \boldsymbol{V}\in\mathbb{R}^{L\times m}_+ \\
& \boldsymbol{\alpha}\in\mathbb{R}^m_+ \ \boldsymbol{q}\in\mathbb{R}^n_+, \ z\in\mathbb{R}.
\end{aligned}
\tag{3.20}
$$

The above formulation is significantly faster than solving (1.2) as we observe in our numerical experiments. Algorithm 2 describes the detail of our algorithm.

Algorithm 2 Computing Approximate Affine Policy

1: **for** $i = 1, \ldots, m$ **do**

2:
$$\boldsymbol{v}_i \in \arg\min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{e}_i \right\}.$$

3: **end for**

4:
$$\boldsymbol{Y} = [\boldsymbol{v}_1 \mid \boldsymbol{v}_2 \mid \ldots \mid \boldsymbol{v}_m].$$

5: Solve the LP :
$$\begin{aligned} z_{\mathsf{Alg}} = \min\ & \boldsymbol{c}^T \boldsymbol{x} + z \\ & z - \boldsymbol{d}^T \boldsymbol{q} \geq \boldsymbol{r}^T \boldsymbol{v} \\ & \boldsymbol{R}^T \boldsymbol{v} \geq \boldsymbol{Y} \cdot \mathsf{diag}(\boldsymbol{\alpha})^T \boldsymbol{d} \\ & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{q} \geq \boldsymbol{V}^T \boldsymbol{r} \\ & \boldsymbol{R}^T \boldsymbol{V} \geq \boldsymbol{I}_m - \boldsymbol{B}\boldsymbol{Y} \cdot \mathsf{diag}(\boldsymbol{\alpha}) \\ & \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{v} \in \mathbb{R}_+^L,\ \boldsymbol{U} \in \mathbb{R}_+^{L \times n},\ \boldsymbol{V} \in \mathbb{R}_+^{L \times m} \\ & \boldsymbol{\alpha} \in \mathbb{R}_+^m\ \boldsymbol{q} \in \mathbb{R}_+^n,\ z \in \mathbb{R}. \end{aligned}$$

6: **return** z_{Alg}.

We would like to note that since our approximate affine solution is based on the construction of affine policy in our analysis, the worst-case approximation bound for our approximate affine solution is also $O(\frac{\log n}{\log \log n})$.

3.5.2 Numerical experiments.

We study the empirical performance of our algorithm for budget of uncertainty sets both from the perspective of computation time and the quality of the solution.

Experimental setup. We use the same test instances as in Ben-Tal et al. [41]. In particular, we choose $n = m$, $\boldsymbol{c} = \boldsymbol{d} = \boldsymbol{e}$ and $\boldsymbol{A} = \boldsymbol{B}$ where $\boldsymbol{B}$ is randomly generated as $\boldsymbol{B} = \boldsymbol{I}_m + \boldsymbol{G}$, where $\boldsymbol{I}_m$ is the identity matrix and $\boldsymbol{G}$ is a random normalized Gaussian, i.e. $G_{ij} = |Y_{ij}|/\sqrt{m}$ where Y_{ij} are i.i.d. standard Gaussians. Let consider the following budget of uncertainty sets:

$$\mathcal{U}_1 = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^m h_i \leq k \right\} \tag{3.21}$$

$$\mathcal{U}_2 = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^m w_i h_i \leq 1 \right\}. \tag{3.22}$$

For our numerical experiments, we choose $k = c\sqrt{m}$ with c a random uniform constant between 1 and 2 for the first uncertainty set $\mathcal{U}_1$. For the second uncertainty set $\mathcal{U}_2$, we choose $\boldsymbol{w}$ a normalized Gaussian vector, i.e., $w_i = |G_i|/\|\boldsymbol{G}\|_2$ where G_i are i.i.d. standard Gaussians. We consider values of m from $m = 10$ to $m = 200$ in increments of 10 and consider 20 instances for each value of m.

We compute the optimal affine solution by solving the LP formulation (A.1). We compute our approximate affine solution returned by Algorithm 2. We denote $z_{\mathsf{Alg}}(\mathcal{U})$ and $z_{\mathsf{Aff}}(\mathcal{U})$ respectively the cost of our affine solution returned by Algorithm 2 and the cost of the optimal affine solution. For each m from $m = 10$ to $m = 200$, we report the average ratio $z_{\mathsf{Alg}}(\mathcal{U})/z_{\mathsf{Aff}}(\mathcal{U})$, the running time of Algorithm 2 in seconds ($T_{\mathsf{Alg}}(s)$) and the running time of the optimal affine policy in seconds ($T_{\mathsf{aff}}(s)$). We present the results of our computational experiments in Table 3.1. The numerical results are obtained using Gurobi 7.0.2 on a 16-core server with 2.93GHz processor and 56GB RAM.

Results. We observe from Table 3.1 that our algorithm is significantly faster than the optimal affine policy. In fact, Algorithm 2 scales very well and the average running time is only a few seconds even for large values of m. On the other hand, computing the optimal affine solution becomes computationally challenging for large values of m. For example, for $m = 100$, the average running time is around 3 minutes for $\mathcal{U}_1$ and more than 11 minutes for $\mathcal{U}_2$. For $m = 200$, the average running time is more than an hour for $\mathcal{U}_1$ and more than 3 hours for $\mathcal{U}_2$. Furthermore, we observe that the gap between our affine solution and the optimal one is under 15%. Moreover, this gap does not increase with the dimension of m, thereby confirming that our affine solution performs well even for large values of m.

m	$T_{\mathsf{aff}}(s)$	$T_{\mathsf{Alg}}(s)$	$z_{\mathsf{Alg}}/z_{\mathsf{Aff}}$
10	0.009	0.022	1.146
20	0.137	0.105	1.111
30	0.304	0.300	1.155
40	1.268	0.692	1.126
50	4.007	1.370	1.120
60	9.461	3.089	1.135
70	17.38	3.417	1.147
80	44.75	5.626	1.103
90	80.20	10.18	1.114
100	153.3	13.23	1.149
200	5137	69.33	1.061

(a) Budget of uncertainty (3.21)

m	$T_{\mathsf{aff}}(s)$	$T_{\mathsf{Alg}}(s)$	$z_{\mathsf{Alg}}/z_{\mathsf{Aff}}$
10	0.011	0.021	1.108
20	0.200	0.110	1.092
30	1.219	0.353	1.103
40	4.887	0.812	1.093
50	17.13	1.388	1.096
60	54.03	2.259	1.086
70	129.7	3.625	1.088
80	248.1	5.069	1.082
90	390.9	6.381	1.080
100	692.9	8.705	1.082
200	**	68.62	**

(b) Budget of uncertainty (3.22)

Table 3.1: Comparison on the performance and computation time of the optimal affine policy and our approximate affine policy. For 20 instances, we compute $z_{\mathsf{Alg}}(\mathcal{U})/z_{\mathsf{Aff}}(\mathcal{U})$ for $\mathcal{U}$ the budget of uncertainty sets (3.21) and (3.22). Here, $T_{\mathsf{Alg}}(s)$ denotes the running time for our approximate affine policy and $T_{\mathsf{aff}}(s)$ denotes the running time for affine policy in seconds. $**$ denotes the cases when we set a time limit of 3 hours. These results are obtained using Gurobi 7.0.2 on a 16-core server with 2.93GHz processor and 56GB RAM.

Remark. The formulation (3.19) provides an approximate affine policy for solving our two-stage adjustable problem under any uncertainty set and not only a single budget of uncertainty set. This approximate affine policy is significantly faster than computing optimal affine policy and has a worst-case approximation bound of $O(\frac{\log n}{\log \log n})$ for single budget of uncertainty set. While our analysis does not provide theoretical guarantees on the performance of this approximate affine policy for general uncertainty sets, it still gives a feasible policy that is significantly faster than computing the optimal affine policy. Moreover, we observe in our extended numerical experiments in Appendix C.3 that the empirical performance of our approximate solution is still good even for intersection of budget of uncertainty sets and the gap is within 20% of the optimal affine policy. However, for general conic sets including ellipsoidal uncertainty sets, the gap between optimal affine policy and our approximate affine policy could be large for some cases (up to a factor of two). We refer the reader to Appendix C.3 for more details.

3.6 General case of recourse matrix

In this section, we consider the two-stage adjustable problem (1.1) with general recourse matrix $\boldsymbol{B}$ where we relax the non-negativity assumption on $\boldsymbol{B}$. In particular, we consider cases where some of the coefficients in $\boldsymbol{B}$ could be negative. We show that in this case, the gap between the optimal affine policy given by (1.2) and the optimal adjustable problem (1.1) could be arbitrary large. Therefore, the non-negativity assumption on the coefficients of $\boldsymbol{B}$ is crucial for affine policies to have a good performance with respect to the optimal adjustable solution.

We consider a two-stage lot-sizing problem to construct a family of instances of (1.1) with general recourse matrix $\boldsymbol{B}$ such that the gap between the optimal adjustable solution and optimal affine policies is unbounded.

Two-stage robust lot-sizing problem. We are given a set of m nodes with pairwise distances d_{ij} between node i and node j. Each node $i \in [m]$ has cost c_i per unit inventory at node i and has a capacity of K_i. Each node i faces an uncertain demand h_i that is realized in the second-stage. In the first-stage, the decision maker needs to decide the inventory levels, x_i for each node $i \in [m]$. We model uncertain demand as an adversarial selection from a pre-specified uncertainty set $\mathcal{U}$ after the adversary observes the first-stage inventory decisions. In the second-stage, the decision maker can make recourse transportation decisions after observing the uncertain demand to satisfy it using the first-stage inventory. The goal is to make the first-stage inventory decisions such that the sum of first-stage inventory costs and the worst case second-stage transportation costs is minimized. This problem has been studied extensively in the literature (see for example Bertsimas and de Ruiter [17]).

We can formulate the above problem in our framework of (1.1) where the recourse matrix $\boldsymbol{B}$ is a network matrix with entries in $\{-1,0,1\}$. The epigraph formulation is the following.

$$
\begin{aligned}
\min_{\boldsymbol{x},z} \; & \sum_{i=1}^{m} c_i x_i + z \\
& z \geq \sum_{i=1}^{m}\sum_{j=1}^{m} d_{ij} y_{ij}(\boldsymbol{h}) \\
& x_i + \sum_{j=1}^{m} y_{ji}(\boldsymbol{h}) - \sum_{j=1}^{m} y_{ij}(\boldsymbol{h}) \geq h_i, \quad \forall i \in [m], \;\; \forall \boldsymbol{h} \in \mathcal{U} \\
& 0 \leq x_i \leq K_i, \quad \forall i \in [m] \\
& \boldsymbol{y}(\boldsymbol{h}) \in \mathbb{R}_+^{m^2}, \quad \forall \boldsymbol{h} \in \mathcal{U}.
\end{aligned}
\tag{3.23}
$$

Family of large gap instances. We consider the following family of instances for the robust lot-sizing problem (3.23). Consider a bipartite network (J_1, J_2) where $|J_1| = |J_2| = m/2$ (m is even). We consider a budget of uncertainty set to model demand uncertainty. The inventory cost c_i, capacity K_i for all $i \in [m]$, distances d_{ij}, $i,j \in [m]$ and the formulation for the uncertainty set are given as follows.

$$
\begin{aligned}
c_i &= \left\{ \begin{array}{ll} 0 & \text{if } i \in J_1 \\ 1 & \text{if } i \in J_2 \end{array} \right\}. \\
K_i &= 1 \qquad \forall i \in [m] \\
d_{ij} &= \left\{ \begin{array}{ll} 0 & \text{if } i \in J_1, j \in J_2 \\ \infty & \text{otherwise.} \end{array} \right\}. \\
\mathcal{U} &= \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} h_i \leq m/2 \right\}.
\end{aligned}
\tag{3.24}
$$

For the above family of instances (3.24), we show that the gap between optimal affine and adjustable policies is unbounded. In particular, we have the following lemma.

Lemma 3.6.1. *For the family of instances* (3.24)*, the optimal adjustable solution is* $z_{\mathsf{AR}}(\mathcal{U}) = 0$ *and the optimal affine solution is* $z_{\mathsf{Aff}}(\mathcal{U}) = m/2 - 1$*. In particular the gap between affine and adjustable policies is unbounded.*

The proof of Lemma 3.6.1 is presented in Appendix C.4. Lemma 3.6.1 shows that the assumption on the non-negativity of the recourse matrix $\boldsymbol{B}$ is necessary and crucial to obtain the theoretical bounds in Table 1.1. Relaxing this assumption can result in an unbounded gap. It is an interesting question to develop approximation algorithms and policies for two-stage robust problem with provable theoretical guarantees when the recourse matrix has negative components, or in particular is a network matrix.

3.7 General case of uncertainty in the constraint matrix.

In this section, we consider the case where the left hand side constraint matrix $\boldsymbol{A}$ in (1.1) depend on the uncertain parameter $\boldsymbol{h}$. We show that even in the case where $\boldsymbol{A}(\boldsymbol{h})$ is an affine function of $\boldsymbol{h}$, the gap between the optimal affine solution and the optimal adjustable solution can be bad and scales linearly with the dimensions of the problem n and m. This shows that our results in Table 1.1 do not extend to the case of uncertainty in the left hand side. Recall the two-stage adjustable problem (1.1) and suppose that the first stage constraint matrix $\boldsymbol{A}$ depends on $\boldsymbol{h}$, i.e.,

$$\begin{aligned} z_{\mathsf{AR}}(\mathcal{U}) = \min_{\boldsymbol{x}} \ & \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \min_{\boldsymbol{y}(\boldsymbol{h})} \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{h}) \\ & \boldsymbol{A}(\boldsymbol{h})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) \ \geq \ \boldsymbol{h}, \quad \forall \boldsymbol{h} \in \mathcal{U} \\ & \boldsymbol{x} \in \mathcal{X} \\ & \boldsymbol{y}(\boldsymbol{h}) \in \mathbb{R}^n_+, \quad \forall \boldsymbol{h} \in \mathcal{U}. \end{aligned} \tag{3.25}$$

Suppose that $\boldsymbol{A}(\boldsymbol{h}) \in \mathbb{R}^{m\times n}$ is an affine function of $\boldsymbol{h}$, i.e.,

$$\boldsymbol{A}(\boldsymbol{h}) = \sum_{i=1}^{m} h_i\boldsymbol{A}_i + \boldsymbol{A}_0$$

where for all $i = 0, \ldots, m$, $\boldsymbol{A}_i \in \mathbb{R}^{m \times n}$.

Family of Large Gap Instances. We consider the following family of instances of problem (3.25),

$$\begin{aligned}
&n = m \qquad \mathcal{X} = \mathbb{R}^m_+ \qquad \boldsymbol{c} = \boldsymbol{0} \qquad \boldsymbol{d} = \boldsymbol{e} \\
&\boldsymbol{A}_0 = \boldsymbol{0} \qquad \boldsymbol{A}_i = (\boldsymbol{e} - \boldsymbol{e}_i)\boldsymbol{e}^T, \ \ \forall i \in [m] \qquad \boldsymbol{B} = \boldsymbol{I}_m \\
&\mathcal{U} = \{\boldsymbol{h} \in [0,1]^m\}.
\end{aligned} \tag{3.26}$$

Note that the uncertainty set $\mathcal{U}$ is a box of uncertainty set which is a special case of the budget of uncertainty set (3.1) with $k = m$. Even under this special case, we show that the gap between the optimal affine solution and the optimal adjustable solution is bad and grows linearly with m. In particular, we have the following lemma.

Lemma 3.7.1. *For the family of instances* (3.26)*, the optimal adjustable solution is* $z_{\mathsf{AR}}(\mathcal{U}) = 1$ *and the optimal affine solution is* $z_{\mathsf{Aff}}(\mathcal{U}) = m/2$*. In particular, the gap between affine and adjustable policies grows linearly with the dimension of the problem m.*

The proof of Lemma 3.7.1 is presented in Appendix C.5. Lemma 3.7.1 shows that our results on the performance of affine policies in Table 1.1 do not extend to the class of problems with left hand side uncertainty where the gap could be as bad as $\Omega(m)$.

Chapter 4: Piecewise affine policies

4.1 Introduction

We consider the problem of designing piecewise affine policies for two-stage adjustable robust linear optimization problems under right-hand side uncertainty. It is well known that a piecewise affine policy is optimal although the number of pieces can be exponentially large. A significant challenge in designing a practical piecewise affine policy is constructing good pieces of the uncertainty set. Here we address this challenge by introducing a new framework in which the uncertainty set is "approximated" by a "dominating" simplex. The corresponding policy is then based on a mapping from the uncertainty set to the simplex. Although our piecewise affine policy has exponentially many pieces, it can be computed efficiently by solving a compact linear program given the dominating simplex. Furthermore, we can find the dominating simplex in a closed form if the uncertainty set satisfies some symmetries and can be computed using a MIP in general. The performance of our policy is significantly better than the affine policy for many important uncertainty sets, such as ellipsoids and norm-balls, both theoretically and numerically. For instance, for hypersphere uncertainty set, our piecewise affine policy can be computed by an LP and gives a $O(m^{1/4})$-approximation whereas the affine policy requires us to solve a second order cone program and has a worst-case performance bound of $O(\sqrt{m})$.

More specifically, recall the two-stage adjustable robust problem (1.1) with covering constraints and uncertain right-hand side:

$$\begin{aligned} z_{\mathsf{AR}}(\mathcal{U}) = \min_{\boldsymbol{x}} \; & \boldsymbol{c}^T \boldsymbol{x} + \max_{\boldsymbol{h} \in \mathcal{U}} \min_{\boldsymbol{y}(\boldsymbol{h})} \boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{h}) \\ & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) \geq \boldsymbol{h}, \quad \forall \boldsymbol{h} \in \mathcal{U} \\ & \boldsymbol{x} \in \mathbb{R}^n_+ \\ & \boldsymbol{y}(\boldsymbol{h}) \in \mathbb{R}^n_+, \quad \forall \boldsymbol{h} \in \mathcal{U}, \end{aligned}$$

As expressed in the above formulation, we refer to $z_{\mathsf{AR}}(\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{A}, \boldsymbol{B}, \mathcal{U})$ as $z_{\mathsf{AR}}(\mathcal{U})$ in this chapter for the sake of simplicity. Moreover, we assume in this chapter that the first-stage decision $\boldsymbol{x}$ belongs to the non-negative orthant, i.e., $\boldsymbol{x} \in \mathcal{X} = \mathbb{R}^n_+$. We assume that the uncertainty set $\mathcal{U}$ satisfies the following assumption.

Assumption 1. $\mathcal{U} \subseteq [0,1]^m$ is convex, full-dimensional with $\boldsymbol{e}_i \in \mathcal{U}$ for all $i = 1, \ldots, m$, and *down-monotone*, i.e., $\boldsymbol{h} \in \mathcal{U}$ and $\boldsymbol{0} \leq \boldsymbol{h}' \leq \boldsymbol{h}$ implies that $\boldsymbol{h}' \in \mathcal{U}$.

We would like to emphasize that the above assumption can be made without loss of generality since we can appropriately scale the uncertainty set, and consider a down-monotone completion, without affecting the two-stage problem (1.1).

Recall that in a Piecewise affine policies (PAP), we consider pieces $\mathcal{U}_i, i \in [k]$ of $\mathcal{U}$ such that $\mathcal{U}_i \subseteq \mathcal{U}$ and $\mathcal{U}$ is covered by the union of all pieces. For each $\mathcal{U}_i$, we have an affine solution $\boldsymbol{y}(\boldsymbol{h})$ where $\boldsymbol{h} \in \mathcal{U}_i$. PAP are significantly more general than static and affine policies. For problem (1.1), with $\mathcal{U}$ being a polytope, a PAP is known to be optimal. However, the number of pieces can be exponentially large. Moreover, finding the optimal pieces is, in general, an intractable task. In fact, Bertsimas and Caramanis [39] prove that it is NP-hard to construct the optimal pieces, even for pieceiwse policies with two pieces, for two-stage robust linear programs. In this chapter, we do not attempt to directly find a partition of $\mathcal{U}$, but we present a tractable new framework to construct piecewise affine policies (PAP) via *dominating* the uncertainty set with a simplex, solving our robust problem over the simplex and recovering a solution over $\mathcal{U}$.

The rest of this chapter is organized as follow. In Section 4.2, we present the new framework for approximating the two-stage adjustable robust problem (1.1) via dominating uncertainty sets and constructing piecewise affine policies. In Section 4.3, we provide improved approximation bounds for (1.1) for scaled permutation invariant sets. We present the case of general uncertainty sets in Section 4.4. In Section 4.5, we present a family of lower-bound instances where our piecewise affine policy has the worst performance bound and finally in Section 4.6, we present a computational study to test our policy and compare it to an affine policy over $\mathcal{U}$.

4.2 A new framework for piecewise affine policies

We present a piecewise affine policy to approximate the two-stage adjustable robust problem (1.1). Our policy is based on approximating the uncertainty set $\mathcal{U}$ with a simple set $\hat{\mathcal{U}}$ such that the adjustable problem (1.1) can be efficiently solved over $\hat{\mathcal{U}}$. In particular, we select $\hat{\mathcal{U}}$ such that it *dominates* $\mathcal{U}$ and it is *close* to $\mathcal{U}$. We make these notions precise with the following definitions.

Definition 4.2.1. (Domination) Given an uncertainty set $\mathcal{U} \subseteq \mathbb{R}^m_+$, $\hat{\mathcal{U}} \subseteq \mathbb{R}^m_+$ dominates $\mathcal{U}$ if for all $\boldsymbol{h} \in \mathcal{U}$, there exists $\hat{\boldsymbol{h}} \in \hat{\mathcal{U}}$ such that $\hat{\boldsymbol{h}} \geq \boldsymbol{h}$.

Definition 4.2.2. (Scaling factor) Given a full-dimensional uncertainty set $\mathcal{U} \subseteq \mathbb{R}^m_+$ and $\hat{\mathcal{U}} \subseteq \mathbb{R}^m_+$ that dominates $\mathcal{U}$. We define the scaling factor $\beta(\mathcal{U},\hat{\mathcal{U}})$ as following

$$\beta(\mathcal{U},\hat{\mathcal{U}}) = \min \left\{ \beta > 0 \mid \hat{\mathcal{U}} \subseteq \beta \cdot \mathcal{U} \right\}.$$

For the sake of simplicity, we denote the scaling factor $\beta(\mathcal{U},\hat{\mathcal{U}})$ by β in the rest of this chapter. The scaling factor always exists since $\mathcal{U}$ is full-dimensional. Moreover, it is greater than one because $\hat{\mathcal{U}}$ dominates $\mathcal{U}$. Note that the dominating set $\hat{\mathcal{U}}$ does not necessarily contain $\mathcal{U}$. We illustrate this in the following example.

Example. Consider the uncertainty set

$$\mathcal{U} = \{\boldsymbol{h} \in \mathbb{R}^m_+ \mid ||\boldsymbol{h}||_2 \leq 1\}. \tag{4.1}$$

which is the intersection of the unit ℓ_2-norm ball and the non-negative orthant. We show later in this chapter (Proposition 4.3.6) that the simplex $\hat{\mathcal{U}}$ dominates $\mathcal{U}$ where

$$\hat{\mathcal{U}} = m^{\frac{1}{4}} \cdot \mathsf{conv}\left(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \frac{1}{\sqrt{m}}\boldsymbol{e}\right). \tag{4.2}$$

Figures 4.1 and 4.2 illustrate the sets $\mathcal{U}$ and $\hat{\mathcal{U}}$ for $m = 3$. Note that $\hat{\mathcal{U}}$ does not contain $\mathcal{U}$ but only dominates $\mathcal{U}$. This is an important property in our framework.

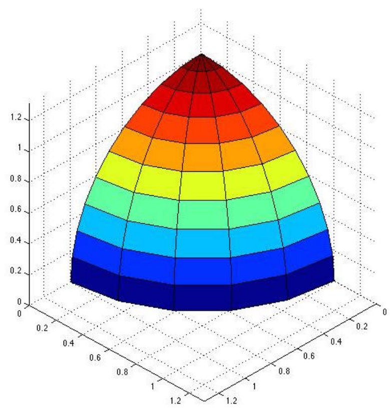

Figure 4.1: The uncertainty set (4.1)

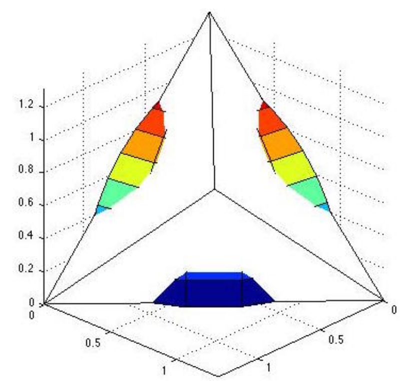

Figure 4.2: The dominating set $\hat{\mathcal{U}}$ (4.2)

The following theorem shows that solving the adjustable problem over the set $\hat{\mathcal{U}}$ gives a β-approximation to the two-stage adjustable robust problem (1.1).

Theorem 4.2.3. *Consider an uncertainty set $\mathcal{U}$ that verifies Assumption* 1 *and $\hat{\mathcal{U}} \subseteq \mathbb{R}^m_+$ that dominates $\mathcal{U}$. Let β be the scaling factor of $(\mathcal{U}, \hat{\mathcal{U}})$. Moreover, let $z_{\mathsf{AR}}(\mathcal{U})$ and $z_{\mathsf{AR}}(\hat{\mathcal{U}})$ be the optimal values for* (1.1) *corresponding to $\mathcal{U}$ and $\hat{\mathcal{U}}$, respectively. Then,*

$$z_{\mathsf{AR}}(\mathcal{U}) \leq z_{\mathsf{AR}}(\hat{\mathcal{U}}) \leq \beta \cdot z_{\mathsf{AR}}(\mathcal{U}).$$

The proof of Theorem 4.2.3 is presented in Appendix E.1.

4.2.1 Choice of $\hat{\mathcal{U}}$

Theorem 4.2.3 provides a new framework for approximating the two-stage adjustable robust problem $\Pi_{\mathsf{AR}}(\mathcal{U})$ (1.1). Note that we require $\hat{\mathcal{U}}$ to be such that it dominates $\mathcal{U}$ and that $\Pi_{\mathsf{AR}}(\hat{\mathcal{U}})$ can be solved efficiently over $\hat{\mathcal{U}}$. In fact, the latter is satisfied if the number of extreme points of $\hat{\mathcal{U}}$ is small and is explicitly given (typically polynomial of m). In our framework, we choose the dominating set to be a simplex of the following form

$$\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\left(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{v}\right), \tag{4.3}$$

for some $\boldsymbol{v} \in \mathcal{U}$. The coefficient β and $\boldsymbol{v} \in \mathcal{U}$ are chosen such that $\hat{\mathcal{U}}$ dominates $\mathcal{U}$. For a given $\hat{\mathcal{U}}$ (i.e., β and $\boldsymbol{v} \in \mathcal{U}$), the adjustable robust problem, $\Pi_{\mathsf{AR}}(\hat{\mathcal{U}})$ (1.1) can be solved efficiently as it can be reduced to the following LP:

$$\begin{aligned} z_{\mathsf{AR}}(\hat{\mathcal{U}}) = \min\ & \boldsymbol{c}^T\boldsymbol{x} + z \\ & z \geq \boldsymbol{d}^T\boldsymbol{y}_i,\ \forall i \in [m+1] \\ & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_i \geq \beta\boldsymbol{e}_i,\ \forall i \in [m] \\ & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}_{m+1} \geq \beta\boldsymbol{v} \\ & \boldsymbol{x} \in \mathbb{R}^n_+,\ \boldsymbol{y}_i \in \mathbb{R}^n_+,\ \forall i \in [m+1]. \end{aligned}$$

4.2.2 Mapping points in $\mathcal{U}$ to dominating points

Consider the following piecewise affine mapping for any $\boldsymbol{h} \in \mathcal{U}$:

$$\forall \boldsymbol{h} \in \mathcal{U}, \qquad \hat{\boldsymbol{h}}(\boldsymbol{h}) = \beta\boldsymbol{v} + (\boldsymbol{h} - \beta\boldsymbol{v})_+. \tag{4.4}$$

We show that this maps any $\boldsymbol{h} \in \mathcal{U}$ to a dominating point contained in the down-monotone completion of $2 \cdot \hat{\mathcal{U}}$. First, the following structural result is needed.

Lemma 4.2.4. (Structural Result) *Consider an uncertainty set $\mathcal{U}$ that verifies Assumption* 1.

a) *Suppose there exists* β *and* $\boldsymbol{v} \in \mathcal{U}$ *such that* $\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{v})$ *dominates* $\mathcal{U}$. *Then,*

$$\frac{1}{\beta} \sum_{i=1}^{m} (h_i - \beta v_i)^+ \leq 1, \; \forall \boldsymbol{h} \in \mathcal{U}. \tag{4.5}$$

b) *Moreover, if there exists* β *and* $\boldsymbol{v} \in \mathcal{U}$ *satisfying* (4.5). *Then,*

$2\beta \cdot \mathsf{conv}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{v})$ *dominates* $\mathcal{U}$.

The proof of Lemma 4.2.4 is presented in Appendix E.2.

The following lemma shows that the mapping in (4.4) maps any $\boldsymbol{h} \in \mathcal{U}$ to a dominating point that belongs to the down-monotone completion of $2 \cdot \hat{\mathcal{U}}$.

Lemma 4.2.5. *For all* $\boldsymbol{h} \in \mathcal{U}$, $\hat{\boldsymbol{h}}(\boldsymbol{h})$ *as defined in* (4.4) *is a dominating point that belongs to the down-monotone completion of* $2 \cdot \hat{\mathcal{U}}$.

Proof. It is clear that $\hat{\boldsymbol{h}}(\boldsymbol{h})$ dominates $\boldsymbol{h}$ because $\hat{\boldsymbol{h}}(\boldsymbol{h}) \geq \beta \boldsymbol{v} + (\boldsymbol{h} - \beta \boldsymbol{v}) = \boldsymbol{h}$. Moreover, for all $\boldsymbol{h} \in \mathcal{U}$, we have

$$\begin{aligned}\hat{\boldsymbol{h}}(\boldsymbol{h}) &= \beta \boldsymbol{v} + \frac{1}{\beta} \sum_{i=1}^{m} (h_i - \beta v_i)^+ \beta \boldsymbol{e}_i \\ &\leq \underbrace{\beta \boldsymbol{v}}_{\in \hat{\mathcal{U}}} + \underbrace{\frac{1}{\beta} \sum_{i=1}^{m} (h_i - \beta v_i)^+ \beta \boldsymbol{e}_i + (1 - \frac{1}{\beta} \sum_{i=1}^{m} (h_i - \beta v_i)^+) \beta \boldsymbol{v}}_{\in \hat{\mathcal{U}}} \in 2 \cdot \hat{\mathcal{U}}\end{aligned}$$

where the inequality

$$1 - \frac{1}{\beta} \sum_{i=1}^{m} (h_i - \beta v_i)^+ \geq 0.$$

follows from part *a*) of Lemma 4.2.4. Therefore, $\hat{\boldsymbol{h}}(\boldsymbol{h})$ belongs to the down-monotone completion of $2 \cdot \hat{\mathcal{U}}$. □

4.2.3 Piecewise affine policy

We construct a *piecewise affine policy over* $\mathcal{U}$ from the optimal solution of $\Pi_{\mathsf{AR}}(\hat{\mathcal{U}})$ based on the piecewise affine mapping in (4.4). Let $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}(\hat{\boldsymbol{h}})$ for $\hat{\boldsymbol{h}} \in \hat{\mathcal{U}}$ be an optimal solution of $\Pi_{\mathsf{AR}}(\hat{\mathcal{U}})$.

Since $\hat{\mathcal{U}}$ is a simplex, we can compute this efficiently.

The piecewise affine policy (PAP)

$$\begin{aligned} \boldsymbol{x} &= 2\hat{\boldsymbol{x}} \\ \boldsymbol{y}(\boldsymbol{h}) &= \frac{1}{\beta}\sum_{i=1}^{m}(h_i - \beta v_i)^+ \, \hat{\boldsymbol{y}}(\beta \boldsymbol{e}_i) + \hat{\boldsymbol{y}}(\beta \boldsymbol{v}), \qquad \forall \boldsymbol{h} \in \mathcal{U}. \end{aligned} \tag{4.6}$$

The following theorem shows that the above PAP gives a 2β-approximation for (1.1).

Theorem 4.2.6. *Consider an uncertainty set $\mathcal{U}$ that verifies Assumption* 1 *and*

$$\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\,(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{v})$$

be a dominating set where $\boldsymbol{v} \in \mathcal{U}$. The piecewise affine solution in (4.6) *is feasible and gives a 2β-approximation for the adjustable robust problem* (1.1).

Proof. First, we show that the policy (4.6) is feasible. We have,

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) &= 2\boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}\left(\frac{1}{\beta}\sum_{i=1}^{m}(h_i - \beta v_i)^+ \, \hat{\boldsymbol{y}}(\beta \boldsymbol{e}_i) + \hat{\boldsymbol{y}}(\beta \boldsymbol{v})\right) \\ &= \left(\boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}\hat{\boldsymbol{y}}(\beta \boldsymbol{v})\right) + \boldsymbol{A}\hat{\boldsymbol{x}} + \frac{1}{\beta}\sum_{i=1}^{m}(h_i - \beta v_i)^+ \, \boldsymbol{B}\hat{\boldsymbol{y}}(\beta \boldsymbol{e}_i) \\ &\geq \left(\boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}\hat{\boldsymbol{y}}(\beta \boldsymbol{v})\right) + \frac{1}{\beta}\sum_{i=1}^{m}(h_i - \beta v_i)^+ \left(\boldsymbol{B}\hat{\boldsymbol{y}}(\beta \boldsymbol{e}_i) + \boldsymbol{A}\hat{\boldsymbol{x}}\right) \\ &\geq \beta \boldsymbol{v} + \sum_{i=1}^{m}(h_i - \beta v_i)^+ \, \boldsymbol{e}_i \\ &\geq \beta \boldsymbol{v} + \sum_{i=1}^{m}(h_i - \beta v_i) \, \boldsymbol{e}_i = \boldsymbol{h}, \end{aligned}$$

where the first inequality follows from part a) of Lemma 4.2.4 and the non-negativity of $\hat{\boldsymbol{x}}$ and $\boldsymbol{A}$. The second inequality follows from the feasibility of $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}(\hat{\boldsymbol{h}})$.

To compute the performance of (4.6), we have for any $\boldsymbol{h} \in \mathcal{U}$,

$$
\begin{aligned}
\boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{h}) &= 2\left(\boldsymbol{c}^T \hat{\boldsymbol{x}} + \boldsymbol{d}^T \left(\frac{1}{2\beta} \sum_{i=1}^{m} (h_i - \beta v_i)^+ \hat{\boldsymbol{y}}(\beta \boldsymbol{e}_i) + \frac{1}{2} \hat{\boldsymbol{y}}(\beta \boldsymbol{v})\right)\right) \\
&\leq 2\left(\boldsymbol{c}^T \hat{\boldsymbol{x}} + \max_{\hat{\boldsymbol{h}} \in \hat{\mathcal{U}}} \boldsymbol{d}^T \hat{\boldsymbol{y}}(\hat{\boldsymbol{h}}) \left(\frac{1}{2\beta} \sum_{i=1}^{m} (h_i - \beta v_i)^+ + \frac{1}{2}\right)\right) \\
&\leq 2\left(\boldsymbol{c}^T \hat{\boldsymbol{x}} + \max_{\hat{\boldsymbol{h}} \in \hat{\mathcal{U}}} \boldsymbol{d}^T \hat{\boldsymbol{y}}(\hat{\boldsymbol{h}})\right) \\
&= 2 \cdot z_{\mathsf{AR}}(\hat{\mathcal{U}}),
\end{aligned}
$$

where the second last inequality follows from part a) of Lemma 4.2.4. From Theorem 4.2.3, $z_{\mathsf{AR}}(\hat{\mathcal{U}}) \leq \beta \cdot z_{\mathsf{AR}}(\mathcal{U})$. Therefore, the cost of the piecewise affine policy for any $\boldsymbol{h} \in \mathcal{U}$

$$
\boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{h}) \leq 2\beta \cdot z_{\mathsf{AR}}(\mathcal{U}),
$$

which implies that the piecewise affine solution (4.6) gives a 2β-approximation for the adjustable robust problem (1.1). □

The above proof shows that it is sufficient to find β and $\boldsymbol{v} \in \mathcal{U}$ satisfying (4.5) in Lemma 4.2.4 to construct a piecewise affine policy that gives a 2β-approximation for (1.1). In particular, we summarize the main result in the following theorem.

Theorem 4.2.7. *Let the uncertainty set* $\mathcal{U}$ *satisfy Assumption* 1. *Consider any* β *and* $\boldsymbol{v} \in \mathcal{U}$ *satisfying* (4.5). *Then, the piecewise affine solution in* (4.6) *gives a* 2β*-approximation for the adjustable robust problem* (1.1).

We would like to note that our piecewise affine policy in not necessarily an optimal piecewise policy. However, for a large class of uncertainty sets, we show that our policy is significantly better than affine policy and can even be computed more efficiently than an affine policy.

4.3 Performance Bounds for Scaled Permutation Invariant Sets

In this section, we present performance bounds of our policy for the class of scaled permutation invariant sets. This class includes ellipsoids, weighted norm-balls, intersection of norm-balls and budget of uncertainty sets. These are widely used uncertainty sets in theory and in practice.

Definition 4.3.1. Scaled Permutation Invariant Sets (SPI)

1. $\mathcal{U}$ is a **permutation invariant set** if $\boldsymbol{x} \in \mathcal{U}$ implies that for any permutation τ of $\{1, 2, \ldots, m\}$, $\boldsymbol{x}^\tau \in \mathcal{U}$ where $x_i^\tau = x_{\tau(i)}$.

2. $\mathcal{U}$ is a **scaled permutation invariant set** if there exists $\boldsymbol{\lambda} \in \mathbb{R}_+^m$ and $\mathcal{V}$ a permutation invariant set such that $\mathcal{U} = \mathsf{diag}(\boldsymbol{\lambda}) \cdot \mathcal{V}$.

For a given SPI set $\mathcal{U}$, it is possible to scale the two-stage adjustable problem (1.1) and get a new problem where the uncertainty set is *permutation invariant* (PI). Indeed, suppose $\mathcal{U} = \mathsf{diag}(\boldsymbol{\lambda}) \cdot \mathcal{V}$ where $\mathcal{V}$ is a permutation invariant set; by multiplying the constraint matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ by $\mathsf{diag}(\boldsymbol{\lambda})^{-1}$, we get a new problem where the uncertainty set now is PI. The performance of our policy is not affected by this scaling. Therefore, without loss of generality, we consider in the rest of this section, the case of permutation invariant uncertainty sets.

We first introduce some structural properties of PI sets. Let $\mathcal{U}$ be PI satisfying Assumption 1. For all $k = 1, \ldots, m$, let

$$\gamma(k) = \frac{1}{k} \cdot \max \left\{ \sum_{i=1}^{k} h_i \;\middle|\; \boldsymbol{h} \in \mathcal{U} \right\}. \tag{4.7}$$

The coefficients, $\gamma(k)$ for all $k = 1, \ldots, m$ affect the geometric structure of $\mathcal{U}$. In particular, we have the following lemma.

Lemma 4.3.2. *Le $\mathcal{U}$ be a permutation invariant set and $\gamma(\cdot)$ be as defined in* (4.7)*. Then,*

$$\gamma(k) \cdot \sum_{i=1}^{k} \boldsymbol{e}_i \in \mathcal{U}, \; \forall k = 1, \ldots, m$$

We present the proof of Lemma 4.3.2 in Appendix E.3. For the sake of simplicity, we denote $\gamma(m)$ by γ in the rest of the chapter. From the above lemma, we know that $\gamma \cdot \boldsymbol{e} \in \mathcal{U}$.

4.3.1 Piecewise affine policy for Permutations Invariant Sets

For any PI set $\mathcal{U}$, we consider the following dominating uncertainty set, $\hat{\mathcal{U}}$ of the form (4.3) with $\boldsymbol{v} = \gamma \boldsymbol{e}$, i.e.,

$$\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\,(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_m, \gamma \boldsymbol{e}) \tag{4.8}$$

where β is the scaling factor guaranteeing that $\hat{\mathcal{U}}$ dominates $\mathcal{U}$. This dominating set $\hat{\mathcal{U}}$ is motivated by the symmetry of the permutation invariant set $\mathcal{U}$. In this section, we show that one can efficiently compute the minimum β such that $\hat{\mathcal{U}}$ in (4.8) dominates $\mathcal{U}$. In particular, we derive an efficiently computable closed-form expression for β, for any PI set $\mathcal{U}$.

From Theorem 4.2.7 we know that to construct a piecewise affine policy with an approximation bound of 2β, it is sufficient to find β such that

$$\frac{1}{\beta} \max_{\boldsymbol{h} \in \mathcal{U}} \sum_{i=1}^{m} (h_i - \beta\gamma)^+ \leq 1 \tag{4.9}$$

and any β implies that $2\beta \cdot \mathsf{conv}\,(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_m, \gamma \boldsymbol{e})$ dominates $\mathcal{U}$ (see Lemma 4.2.4b). Finding the minimum β that satisfies (4.9) requires solving:

$$\min \left\{ \beta \geq 1 \;\middle|\; \frac{1}{\beta} \max_{\boldsymbol{h} \in \mathcal{U}} \sum_{i=1}^{m} (h_i - \beta\gamma)^+ \leq 1 \right\}. \tag{4.10}$$

The following lemma characterizes the structure of the optimal solution for the maximization problem in (4.9) for a fixed β.

Lemma 4.3.3. *Consider the maximization problem in* (4.9) *for a fixed β. There exists an optimal solution $\boldsymbol{h}^*$ such that*

$$\boldsymbol{h}^* = \gamma(k) \cdot \sum_{i=1}^{k} \boldsymbol{e}_i,$$

for some $k = 1, \ldots, m$.

We present the proof of Lemma 4.3.3 in Appendix E.4. The following lemma characterizes the optimal β for (4.10).

Lemma 4.3.4. *Let $\mathcal{U}$ be a permutation invariant uncertainty set satisfying Assumption* 1. *Then the optimal solution for* (4.10) *is given by*

$$\beta = \max_{k=1,\ldots,m} \frac{\gamma(k)}{\gamma + \frac{1}{k}}. \tag{4.11}$$

Proof. Using Lemma 4.3.3, we can reformulate (4.10) as follows.

$$\min \left\{ \beta \geq 1 \;\middle|\; \frac{1}{\beta} \max_{k=1,\ldots,m} \sum_{i=1}^{k} (\gamma(k) - \beta\gamma) \leq 1 \right\},$$

i.e.,

$$\min \left\{ \beta \geq 1 \;\middle|\; \beta \geq \frac{\gamma(k)}{\gamma + \frac{1}{k}}, \;\; \forall k = 1, \ldots, m \right\}.$$

Therefore,

$$\beta = \max_{k=1,\ldots,m} \frac{\gamma(k)}{\gamma + \frac{1}{k}}.$$

□

The above lemma computes the minimum β that satisfies (4.9). Therefore, from Theorem 4.2.7, we have the following theorem.

Theorem 4.3.5. *Let $\mathcal{U}$ be a permutation invariant set satisfying Assumption 1. Let $\gamma = \gamma(m)$ be as defined in* (4.7) *and β be as defined in* (4.11)*, and*

$$\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\left(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \gamma\boldsymbol{e}\right).$$

Let $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}(\hat{\boldsymbol{h}})$ for $\hat{\boldsymbol{h}} \in \hat{\mathcal{U}}$ be an optimal solution for $\Pi_{\mathsf{AR}}(\hat{\mathcal{U}})$ (1.1)*. Then the following piecewise* *affine solution*

$$\begin{aligned} \boldsymbol{x} &= 2\hat{\boldsymbol{x}} \\ \boldsymbol{y}(\boldsymbol{h}) &= \frac{1}{\beta}\sum_{i=1}^{m}(h_i - \beta\gamma)^+\,\hat{\boldsymbol{y}}(\beta\boldsymbol{e}_i) + \hat{\boldsymbol{y}}(\beta\gamma\boldsymbol{e}) \qquad \forall \boldsymbol{h} \in \mathcal{U}, \end{aligned} \tag{4.12}$$

gives a 2β*-approximation for* (1.1). *Moreover, the set* $2 \cdot \hat{\mathcal{U}}$ *dominates* $\mathcal{U}$.

The last claim that $2 \cdot \hat{\mathcal{U}}$ dominates $\mathcal{U}$ is a straightforward consequence of part(b) of Lemma 4.2.4.

As a consequence of Theorem 4.3.5, for any permutation invariant uncertainty set, $\mathcal{U}$, we can compute the piecewise-affine policy for (1.1) efficiently. In fact, for many cases, even more efficiently than an affine policy.

4.3.2 Examples

We present the approximation bounds for several permutation invariant uncertainty sets that are commonly used in the literature and in practice, including norm balls, intersection of norm balls and budget of uncertainty sets. In particular, it follows that for these sets, the performance bounds of our piecewise affine policy are significantly better than the best known performance bounds for affine policy.

Propostion 4.3.6. (Hypersphere) Consider the uncertainty set $\mathcal{U} = \{\boldsymbol{h} \in \mathbb{R}_+^m \mid ||\boldsymbol{h}||_2 \le 1\}$ which is the intersection of the unit hypersphere and the nonnegative orthant. Then,

$$\hat{\mathcal{U}} = m^{\frac{1}{4}} \cdot \mathrm{conv}\left(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_m, \frac{1}{\sqrt{m}}\boldsymbol{e}\right),$$

dominates $\mathcal{U}$ and our piecewise affine solution (4.12) gives $O(m^{\frac{1}{4}})$ approximation to (1.1).

Proof. We have for $k = 1, \ldots, m$,

$$\gamma(k) = \frac{1}{k} \cdot \max\left\{\sum_{i=1}^{k} h_i \mid \boldsymbol{h} \in \mathcal{U}\right\} = \frac{1}{\sqrt{k}}.$$

In particular, $\gamma = \frac{1}{\sqrt{m}}$. From Lemma 4.3.4 we get,

$$\beta = \max_{k=1,\ldots,m} \frac{\gamma(k)}{\gamma(m) + \frac{1}{k}}$$
$$= \max_{k=1,\ldots,m} \frac{\frac{1}{\sqrt{k}}}{\frac{1}{\sqrt{m}} + \frac{1}{k}}.$$

The maximum of this problem occurs for $k = \sqrt{m}$. Then, $\beta = \frac{m^{\frac{1}{4}}}{2}$. We conclude from Theorem 4.3.5 that $\hat{\mathcal{U}}$ dominates $\mathcal{U}$ and our piecewise affine policy gives $O(m^{\frac{1}{4}})$ approximation to the adjustable problem (1.1). □

Remark. Consider the following ellipsoid uncertainty set

$$\left\{ \boldsymbol{h} \geq \boldsymbol{0} \;\middle|\; \sum_{i=1}^{m} r_i h_i^2 \leq 1 \right\}. \tag{4.13}$$

This is widely used to model uncertainty in practice and is just a diagonal scaling of the hypersphere uncertainty set. As we mention before, the performance of our policy is not affected by scaling. Hence, our piecewise affine policy gives an $O(m^{\frac{1}{4}})$-approximation to the adjustable problem (1.1) for ellipsoid uncertainty sets (4.13) similar to hypersphere. We analyze the case of more general ellipsoids in Proposition 4.3.9.

Propostion 4.3.7. (p-norm ball). Consider the p-norm ball uncertainty set

$$\mathcal{U} = \left\{ \boldsymbol{h} \in \mathbb{R}_+^m \mid \|\boldsymbol{h}\|_p \leq 1 \right\}$$

where $p \geq 1$. Then

$$\hat{\mathcal{U}} = 2\beta \cdot \mathsf{conv}\left(\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_m}, m^{-\frac{1}{p}} \boldsymbol{e} \right)$$

dominates $\mathcal{U}$ with

$$\beta = \frac{1}{p}(p-1)^{\frac{p-1}{p}} \cdot m^{\frac{p-1}{p^2}} = O(m^{\frac{p-1}{p^2}}).$$

Our piecewise affine solution (4.12) gives $O(m^{\frac{p-1}{p^2}})$ approximation to (1.1).

Proof. We have for $k = 1, \ldots, m$,

$$\gamma(k) = \frac{1}{k} \cdot \max\left\{\sum_{i=1}^{k} h_i \mid \boldsymbol{h} \in \mathcal{U}\right\} = k^{\frac{-1}{p}}.$$

In particular, $\gamma = m^{\frac{-1}{p}}$. From Lemma 4.3.4 we get,

$$\begin{aligned}\beta &= \max_{k=1,\ldots,m} \frac{\gamma(k)}{\gamma(m) + \frac{1}{k}} \\ &= \max_{k=1,\ldots,m} \frac{k^{\frac{-1}{p}}}{m^{\frac{-1}{p}} + \frac{1}{k}} \\ &= \frac{1}{p}(p-1)^{\frac{p-1}{p}} \cdot m^{\frac{p-1}{p^2}} = O\left(m^{\frac{p-1}{p^2}}\right).\end{aligned}$$

We conclude from Theorem 4.3.5 that $\hat{\mathcal{U}}$ dominates $\mathcal{U}$ and our piecewise affine policy gives $O(m^{\frac{p-1}{p^2}})$ approximation to the adjustable problem (1.1). □

Propostion 4.3.8. (Intersection of two norm balls) Consider $\mathcal{U}$ the intersection of the norm balls
$\mathcal{U}_1 = \left\{\boldsymbol{h} \in \mathbb{R}_+^m \mid \|\boldsymbol{h}\|_p \leq 1\right\}$ and
$\mathcal{U}_2 = \left\{\boldsymbol{h} \in \mathbb{R}_+^m \mid \|\boldsymbol{h}\|_q \leq r\right\}$ where $p > q \geq 1$ and $m^{\frac{1}{q}-\frac{1}{p}} \geq r \geq 1$. Then,

$$\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\left(\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_m}, \left(rm^{-\frac{1}{q}}\right)\boldsymbol{e}\right),$$

where

$$\beta = \min(\beta_1, \beta_2),\ \beta_1 = r^{\frac{1-p}{p}} m^{\frac{p-1}{pq}},\text{ and } \beta_2 = r^{\frac{1}{q}} m^{\frac{q-1}{q^2}}.$$

Our piecewise affine solution (4.12) gives a 2β approximation to (1.1).

Proof. To prove that $\hat{\mathcal{U}}$ dominates $\mathcal{U}_1 \cap \mathcal{U}_2$, it is sufficient to consider $\boldsymbol{h}$ in the boundary of $\mathcal{U}_1$ or

$\mathcal{U}_2$ and find $\alpha_1, \alpha_2, \ldots, \alpha_{m+1} \geq 0$ with $\alpha_1 + \ldots + \alpha_{m+1} = 1$ such that for all $i \in [m]$,

$$h_i \leq \beta \left(\alpha_i + rm^{-\frac{1}{q}} \alpha_{m+1} \right).$$

Case 1: $\beta = \beta_1$.

Let $\boldsymbol{h} \in \mathcal{U}_1$ such that $\|\boldsymbol{h}\|_p = 1$, we take $\alpha_i = \frac{h_i^p}{p}$ for $i \in [m]$ and $\alpha_{m+1} = \frac{p-1}{p}$. First, we have $\sum_{i=1}^{m+1} \alpha_i = 1$ and for all $i \in [m]$,

$$\begin{aligned} \beta \left(\alpha_i + rm^{-\frac{1}{q}} \alpha_{m+1} \right) &= \beta_1 \left(\frac{h_i^p}{p} + \frac{p-1}{p} rm^{-\frac{1}{q}} \right) \\ &\geq \beta_1 \left(h_i^p \right)^{\frac{1}{p}} \left(rm^{-\frac{1}{q}} \right)^{\frac{p-1}{p}} = h_i, \end{aligned}$$

where the inequality follows from the weighted inequality of arithmetic and geometric means (known as Weighted AM-GM inequality). Therefore $\hat{\mathcal{U}}$ dominates $\mathcal{U}_1 \cap \mathcal{U}_2$.

Case 2: $\beta = \beta_2$.

Let $\boldsymbol{h} \in \mathcal{U}_2$ such that $\|\boldsymbol{h}\|_q = r$, we take $\alpha_i = \frac{h_i^q}{r^q q}$ for $i \in [m]$ and $\alpha_{m+1} = \frac{q-1}{q}$. First, we have $\sum_{i=1}^{m+1} \alpha_i = 1$ and for all $i \in [m]$,

$$\begin{aligned} \beta \left(\alpha_i + rm^{-\frac{1}{q}} \alpha_{m+1} \right) &= \beta_2 \left(\frac{h_i^q}{r^q q} + \frac{q-1}{q} rm^{-\frac{1}{q}} \right) \\ &\geq \beta_2 \left(\frac{h_i^q}{r^q} \right)^{\frac{1}{q}} \left(rm^{-\frac{1}{q}} \right)^{\frac{q-1}{q}} = h_i, \end{aligned}$$

where the inequality followed from the weighted AM-GM inequality. Therefore, $\hat{\mathcal{U}}$ dominates $\mathcal{U}_1 \cap \mathcal{U}_2$. □

We also consider a permutation invariant uncertainty set that is the intersection of an ellipsoid and the non-negative orthant , i.e.,

$$\mathcal{U} = \left\{ \boldsymbol{h} \in \mathbb{R}_+^m \mid \boldsymbol{h}^T \boldsymbol{\Sigma} \boldsymbol{h} \le 1 \right\} \tag{4.14}$$

where $\boldsymbol{\Sigma} \succeq \boldsymbol{0}$. For $\mathcal{U}$ to be a permutation invariant set satisfying Assumption 1, $\boldsymbol{\Sigma}$ must be of the following form

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & a & \dots & a \\ a & 1 & \dots & a \\ \vdots & \vdots & \ddots & \vdots \\ a & a & \dots & 1 \end{pmatrix} \tag{4.15}$$

where $0 \le a \le 1$.

Propostion 4.3.9. (Permutation invariant ellipsoid) Consider the uncertainty set $\mathcal{U}$ defined in (4.14) where $\boldsymbol{\Sigma}$ is defined in (4.15). Then

$$\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\left(\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_m, \gamma \boldsymbol{e}\right),$$

dominates $\mathcal{U}$ with

$$\beta = \left(\frac{a}{2} + \frac{(1-a)^{\frac{1}{2}}}{(am^2 + (1-a)m)^{\frac{1}{4}}} \right)^{-1} = O\left(m^{\frac{2}{5}}\right)$$

and

$$\gamma = \frac{1}{\sqrt{(am^2 + (1-a)m)}}.$$

Our piecewise affine policy (4.12) gives $O\left(m^{\frac{2}{5}}\right)$ approximation to the adjustable robust problem (1.1).

The proof of Proposition 4.3.9 is presented in Appendix E.5.

Propostion 4.3.10. (Budget of uncertainty set) Consider the budget of uncertainty set

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \mid \sum_{i=1}^m h_i \le k \right\}. \tag{4.16}$$

Then,

$$\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\left(\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_m}, \frac{k}{m}\boldsymbol{e}\right)$$

where $\beta = \min\left(k, \frac{m}{k}\right)$. In particular, our piecewise affine policy (4.12) gives 2β approximation to the adjustable problem (1.1).

The proof of Proposition 4.3.10 is presented in Appendix E.6.

4.3.3 Comparison to affine policy

Table 1.2 summarizes the performance bounds for our piecewise affine policy and the best known performance bounds in the literature for affine policies [40]. As can be seen, our piecewise affine policy performs significantly better than the known bounds for affine policy for many interesting sets, including hypersphere, ellipsoid and norm-balls. For instance, our policy gives $O(m^{\frac{1}{4}})$-approximation for the hypersphere and $O(m^{\frac{p-1}{p^2}})$-approximation for the p-norm ball, while affine policy gives $O(m^{\frac{1}{2}})$-approximation for hypersphere and $O(m^{\frac{1}{p}})$-approximation for the p-norm ball [40], respectively. However, as we mentioned before, our policy is not a generalization of affine policies and, in fact, affine policies may perform better for certain uncertainty sets. However, we present a family of examples where an optimal affine policy gives an $\Omega(\sqrt{m})$-approximation, while our policy is *near-optimal* for the adjustable robust problem (1.1). In particular, we consider the following instance motivated from the worst-case examples of affine policy in [18] and [20].

$$\begin{aligned}
& n = m, \;\; r = \lceil m - \sqrt{m} \rceil, \;\; N = \binom{m}{r} \\
& B_{ij} = \left\{ \begin{array}{ll} 1 & \text{if } \; i = j \\ \frac{1}{\sqrt{m}} & \text{if } \; i \neq j \end{array} \right\}. \\
& \boldsymbol{A} = \boldsymbol{B}, \;\; \boldsymbol{c} = \frac{1}{15}\boldsymbol{e}, \;\; \boldsymbol{d} = \boldsymbol{e} \\
& \mathcal{U} = \mathsf{conv}\left(\boldsymbol{0}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_N\right) \\
& \text{where } \boldsymbol{\nu}_1 = \frac{1}{\sqrt{m}} \cdot [\underbrace{1, \ldots, 1}_{r}, 0 \ldots, 0];
\end{aligned} \tag{4.17}$$

$\boldsymbol{v}_1$ has exactly r non-zero coordinates, each equal to $\frac{1}{\sqrt{m}}$. The extreme points $\boldsymbol{v}_i$ of v_1, are permutations of the non-zero coordinates of $\boldsymbol{v}_1$. Therefore, $\mathcal{U}$ has exactly $\binom{m}{r} + m + 1$ extreme points.

Lemma 4.3.11. *Our piecewise affine policy* (4.6) *gives an* $O(1 + \frac{1}{\sqrt{m}})$*-approximation for the adjustable robust problem* (1.1) *for instance* (4.17).

We can prove Lemma 4.3.11 by constructing a dominating set within a scaling factor $O(1+\frac{1}{\sqrt{m}})$ from $\mathcal{U}$. We present the complete proof of Lemma 4.3.11 in Appendix E.7.

Lemma 4.3.12. *Affine policy gives an* $\Omega(\sqrt{m})$*-approximation for the adjustable robust problem* (1.1) *for instance* (4.17). *Moreover, for any optimal affine solution, the cost of the first-stage solution* $\boldsymbol{x}^*_{\mathsf{Aff}}$ *is* $\Omega(\sqrt{m})$ *away from the optimal adjustable problem* (1.1)*, i.e.* $\boldsymbol{c}^T\boldsymbol{x}^*_{\mathsf{Aff}} = \Omega(m^{1/2}) \cdot z_{\mathsf{AR}}(\mathcal{U})$.

We present the proof of Lemma 4.3.12 in Appendix E.8. From Lemma 4.3.12 and 4.3.11, we conclude that our policy is near-optimal whereas affine policy is $\Omega(\sqrt{m})$ away from the optimal adjustable solution for the instance (4.17). Hence our policy provides a significant improvement. We would like to note that since $\hat{\mathcal{U}}$ is a simplex, an affine policy is optimal for $\Pi_{\mathsf{AR}}(\hat{\mathcal{U}})$. In particular, we have the following

$$z_{\mathsf{AR}}(\mathcal{U}) \leq z_{\mathsf{AR}}(\hat{\mathcal{U}}) = z_{\mathsf{Aff}}(\hat{\mathcal{U}}) \leq O\left(1 + \frac{1}{\sqrt{m}}\right) \cdot z_{\mathsf{AR}}(\mathcal{U}),$$

where the first inequality follows as $\hat{\mathcal{U}}$ dominates $\mathcal{U}$ and the last inequality follows from Lemma 4.3.11. Moreover, from Lemma 4.3.12, we know that for instance (4.17),

$$z_{\mathsf{Aff}}(\mathcal{U}) = \Omega(\sqrt{m}) \cdot z_{\mathsf{AR}}(\mathcal{U}).$$

Therefore,

$$z_{\mathsf{Aff}}(\mathcal{U}) = \Omega(\sqrt{m}) \cdot z_{\mathsf{Aff}}(\hat{\mathcal{U}}),$$

which is quite surprising since $\hat{\mathcal{U}}$ dominates $\mathcal{U}$. We would like to emphasize that $\hat{\mathcal{U}}$ only dominates $\mathcal{U}$ and does not contain it and this is crucial to get a significant improvement for our piecewise affine policy constructed through the dominating set.

Comparison to re-solving policy: In many applications, a practical implementation of affine policy only implements the first stage solution $\boldsymbol{x}^*_{\mathsf{Aff}}$ and re-solve (or recompute) the second-stage solution once the uncertainty is realized. The performance of such a re-solving policy is at least as good as affine policy and in many cases significantly better. Lemma 4.3.12 shows that for instance (4.17), such a re-solving policy is $\Omega(\sqrt{m})$ away from the optimal adjustable policy whereas we show in Lemma 4.3.11 that our piecewise affine policy is near-optimal. Hence, our piecewise affine policy for instance (4.17) is performing significantly better not only than affine policy but also the re-solving policy.

4.4 General uncertainty set

In this section, we consider the case of general uncertainty sets. The main challenge in our framework of constructing the piecewise affine policy is the choice of the dominating simplex, $\hat{\mathcal{U}}$. More specifically, the choice of β and $\boldsymbol{v} \in \mathcal{U}$ such that $\beta \cdot \mathsf{conv}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{v})$ dominates $\mathcal{U}$. For a permutation invariant set, $\mathcal{U}$, we choose $\boldsymbol{v} = \gamma \boldsymbol{e}$ and we can efficiently find β using Lemma 4.3.4 to construct the dominating set. However, this does not extend to general sets and we need a new procedure to find those parameters.

Theorem 4.2.7 shows that to construct a good piecewise affine policy over $\mathcal{U}$, it is sufficient to find β and $\boldsymbol{v} \in \mathcal{U}$ such that for all $\boldsymbol{h} \in \mathcal{U}$

$$\frac{1}{\beta} \sum_{i=1}^{m} (h_i - \beta v_i)^+ \leq 1. \tag{4.18}$$

In this section, we present an iterative algorithm to find such β and $\boldsymbol{v} \in \mathcal{U}$ satisfying (4.18). In each iteration t, the algorithm maintains a candidate solution, β^t and $\boldsymbol{v}^t \in \mathcal{U}$. Let $\boldsymbol{u}^t = \beta^t \cdot v^t$. The algorithm solves the following maximization problem:

$$\max_{\boldsymbol{h} \in \mathcal{U}} \sum_{i=1}^{m} \left(h_i - u_i^t\right)^+ \tag{4.19}$$

Algorithm 3 Computing β and $\boldsymbol{v}$ for general uncertainty sets

Initialize $t = 0,\ \boldsymbol{u}^0 = 0$
while $\left\{\max\limits_{\boldsymbol{h}\in\mathcal{U}} \sum_{i=1}^m \left(h_i - u_i^t\right)^+ > t\right\}$ **do**
  $\boldsymbol{h}^t \in \operatorname{argmax}\limits_{\boldsymbol{h}\in\mathcal{U}} \sum_{i=1}^m \left(h_i - u_i^t\right)^+$
  for $i = 1, \ldots, m$ **do**
    if $u_i^t = 1$ **then** $h_i^t = 0$
    end if
    $u_i^{t+1} = \min(1, u_i^t + h_i^t)$
  end for
  $t = t + 1$
end while
return $\beta = t,\ \boldsymbol{v} = \frac{\boldsymbol{u}^t}{\beta}$.

The algorithm stops if the optimal value is at most β^t in which case, Condition (4.18) is verified for all $h \in \mathcal{U}$. Otherwise, let $\boldsymbol{h}^t$ be an optimal solution of problem (4.19). The current solutions are updated as follows:

$$\beta^{t+1} = \beta^t + 1$$

$$u_i^{t+1} = \min\left(1, u_i^t + h_i^t\right).$$

This corresponds to updating $\boldsymbol{v}^{t+1} = \frac{1}{\beta^{t+1}} \cdot \boldsymbol{u}^{t+1}$. Algorithm 3 presents the steps in detail.

The number of β-iterations is finite since $\mathcal{U}$ is compact. The following theorem shows that $\boldsymbol{v}$ returned by the algorithm belongs to $\mathcal{U}$ and the corresponding piecewise affine policy is a $O(\sqrt{m})$-approximation for the adjustable problem (1.1).

Theorem 4.4.1. *Suppose Algorithm 3 returns β, $\boldsymbol{v}$. Then $\boldsymbol{v} \in \mathcal{U}$. Furthermore, the piecewise affine policy* (4.6) *with parameters β and $\boldsymbol{v}$ gives a $O(\sqrt{m})$-approximation for the adjustable problem* (1.1).

Proof. Suppose Algorithm 3 returns $\beta, \boldsymbol{v}$. Note that β is the number of iterations in Algorithm 3. First, we have

$$\boldsymbol{u}^\beta \leq \sum_{t=0}^{\beta-1} \boldsymbol{h}^t.$$

Moreover $\frac{1}{\beta} \cdot \sum_{t=0}^{\beta-1} \boldsymbol{h}^t \in \mathcal{U}$ since $\mathcal{U}$ is convex. Therefore $\boldsymbol{v} = \frac{\boldsymbol{u}^\beta}{\beta} \in \mathcal{U}$ by down-monotonicity of

$\mathcal{U}$.

Let us prove that $\beta = O(\sqrt{m})$. First, note that, when we set $h_i^t = 0$ for $u_i^t = 1$, the objective of the maximization problem in the algorithm does not change and $\boldsymbol{h}^t$ still belongs to $\mathcal{U}$ by down-monotonicity. Then, for any $t = 0, \ldots, \beta - 1$

$$\sum_{i=1}^{m} \left(h_i^t - u_i^t\right)^+ > t.$$

Moreover, $h_i^t \geq 0$ and $u_i^t \geq 0$, hence $h_i^t \geq (h_i^t - u_i^t)^+$ and therefore for all $t = 0, \ldots, \beta - 1$

$$\sum_{i=1}^{m} h_i^t > t.$$

Then,

$$\sum_{t=0}^{\beta-1} \sum_{i=1}^{m} h_i^t > \sum_{t=0}^{\beta-1} t = \frac{1}{2}\beta(\beta - 1). \tag{4.20}$$

Note that, if $u_i^t = 1$ at some iteration t, then $h_i^{t'} = 0$ for any $t' \geq t$. Hence, for any $i \in [m]$,

$$\sum_{t=0}^{\beta-1} h_i^t \leq u_i^\beta + 1 \leq 2. \tag{4.21}$$

Hence, from (4.20) and from (4.21) we get, $2m > \frac{1}{2}\beta(\beta - 1)$, i.e., $\beta \cdot (\beta - 1) \leq 4m$, which implies, $\beta = O(\sqrt{m})$. □

We note that the maximization problem (4.19) that Algorithm 3 solves in each iteration t is not a convex optimization problem. However, (4.19) can be formulated as the following MIP:

$$
\begin{aligned}
\max\ & \sum_{i=1}^{m} z_i \\
& z_i \leq (h_i - u_i^t) + (1 - x_i) \ \ \forall i \in [m], \\
& z_i \leq x_i \ \ \forall i \in [m] \\
& z_i \geq 0, \ \ \forall i \in [m] \\
& x_i \in \{0,1\} \ \ \forall i \in [m] \\
& \boldsymbol{h} \in \mathcal{U}.
\end{aligned}
\tag{4.22}
$$

Therefore, for general uncertainty set $\mathcal{U}$, the procedure to find β and $\boldsymbol{v} \in \mathcal{U}$ is computationally more challenging than for the case of permutation invariant sets.

Remark. Since the computation of β and $\boldsymbol{v}$ depends only on $\mathcal{U}$, and not on the problem parameters (i.e., the parameters $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{c}$ and $\boldsymbol{d}$), one can compute them offline and then use them to efficiently construct a good piecewise affine policy.

Connection to Bertsimas and Goyal [18]. We would like to note that Algorithm 3 is quite analogous to the explicit construction of good affine policies in [18]. The analysis of the $O(\sqrt{m})$-approximation bound for affine policies is based on the following projection result (which is a restatement of Lemma 8 and Lemma 9 in [18]).

Theorem 4.4.2. [Bertsimas and Goyal 2011] *Consider any uncertainty set $\mathcal{U}$ satisfying Assumption* 1. *There exists $\beta \leq \sqrt{m}$, $\boldsymbol{v} \in \mathcal{U}$ such that*

$$
\sum_{j:\beta v_j < 1} h_j \leq \beta, \ \forall \boldsymbol{h} \in \mathcal{U}.
$$

Suppose $J = \{j \mid \beta v_j < 1\}$. The affine solution in [18] covers $\beta \boldsymbol{v}$ using the static component and the components J using a linear solution. The linear solution does not exploit the coverage of βv_i for $i \in J$ from the static solution. The approximation factor is $O(\beta)$ since for all $\boldsymbol{h} \in \mathcal{U}$,

$\sum_{j \in J} h_j \le \beta$.

Our piecewise affine solution given by Algorithm 3 finds analogous β, $\boldsymbol{v} \in \mathcal{U}$ such that

$$\sum_{i=1}^{m} (h_i - \beta v_i)_+ \le \beta, \; \forall \boldsymbol{h} \in \mathcal{U}.$$

In the piecewise affine solution, the static component covers $\beta \boldsymbol{v}$ and the remaining part $(\boldsymbol{h} - \beta \boldsymbol{v})_+$ is covered by a piecewise-linear function that exploits the coverage of $\beta \boldsymbol{v}$. This allows us to improve significantly as compared to the affine policy for a large family of uncertainty sets. We would like to note again that our policy is not necessarily an optimal one and there can be examples where affine policy is better than our policy.

4.5 A worst case example for the domination policy

From Theorem 4.4.1, we know that our piecewise affine policy gives an $O(\sqrt{m})$-approximation for the adjustable robust problem (1.1). In this section, we show that this bound is tight for the following budget of uncertainty set:

$$\mathcal{U} = \left\{ \boldsymbol{h} \in \mathbb{R}_+^m \;\middle|\; \sum_{i=1}^{m} h_i = \sqrt{m}, \; 0 \le h_i \le 1 \;\; \forall i \in [m] \right\}. \tag{4.23}$$

We show that our dominating simplex based piecewise affine policy gives an $\Omega(\sqrt{m})$-approximation to the adjustable robust problem (1.1). The lower bound of $\Omega(\sqrt{m})$ holds even when we consider more general dominating sets than simplex. We show that for any $\epsilon > 0$, there is no polynomial number of points in $\mathcal{U}$ such that the convex hull of those points scaled by $m^{\frac{1}{2}-\epsilon}$ dominates $\mathcal{U}$. In particular, we have the following theorem.

Theorem 4.5.1. *Given any* $0 < \epsilon < 1/2$, *and* $k \in \mathbb{N}$, *consider the budget of uncertainty set,* $\mathcal{U}$ (4.23) *with m sufficiently large. Let* $P(m) \le m^k$. *Then for any* $z_1, z_2, \ldots z_{P(m)} \in \mathcal{U}$, *the set*

$$\hat{\mathcal{U}} = m^{\frac{1}{2}-\epsilon} \cdot \mathsf{conv}\left(z_1, z_2, \ldots z_{P(m))}\right),$$

does not dominate $\mathcal{U}$.

Proof. Suppose for a sake of contradiction that there exists $z_1, z_2, \ldots, z_{P(m)} \in \mathcal{U}$ such that $\hat{\mathcal{U}} = m^{\frac{1}{2}-\epsilon} \cdot \mathsf{conv}\,(z_1, z_2, \ldots z_{P(m)})$ dominates $\mathcal{U}$.

By Caratheodory's theorem, we know that any point in $\mathcal{U}$ can be expressed as a convex combination of at most $m+1$ extreme points of $\mathcal{U}$. Therefore

$$\hat{\mathcal{U}} \subseteq m^{\frac{1}{2}-\epsilon} \cdot \mathsf{conv}\left(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{Q(m))}\right),$$

where $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{Q(m)}$ are extreme points of $\mathcal{U}$ and

$$Q(m) \leq (m+1) \cdot P(m) = O(m^{k+1}).$$

Consider any $I \subseteq \{1, 2, \ldots, m\}$ such that $|I| = \sqrt{m}$. Let $\boldsymbol{h}$ be an extreme point of $\mathcal{U}$ corresponding to I, i.e., $h_i = 1$ if $i \in I$ and $h_i = 0$ otherwise. Since we assume that $\hat{\mathcal{U}}$ dominates $\mathcal{U}$, there exists $\hat{\boldsymbol{h}} \in \hat{\mathcal{U}}$ such that $\boldsymbol{h} \leq \hat{\boldsymbol{h}}$. Let

$$\hat{\boldsymbol{h}} = m^{\frac{1}{2}-\epsilon} \sum_{j=1}^{Q(m)} \alpha_j \boldsymbol{y}_j,$$

where $\sum_{j=1}^{Q(m)} \alpha_j = 1$ and $\alpha_j \geq 0$ for all $j = 1, 2, \ldots, Q(m)$. We have

$$1 = h_i \leq \hat{h}_i \ \ \forall i \in I$$

i.e.

$$1 \leq m^{\frac{1}{2}-\epsilon} \sum_{j=1}^{Q(m)} \alpha_j y_{ji}, \ \forall i \in I.$$

Summing over $i \in I$, we have,

$$\sqrt{m} = |I| \leq m^{\frac{1}{2}-\epsilon} \sum_{i \in I} \sum_{j=1}^{Q(m)} \alpha_j y_{ji}.$$

Therefore,

$$
\begin{aligned}
m^{\epsilon} &\leq \sum_{j=1}^{Q(m)} \alpha_j \sum_{i\in I} y_{ji}, \\
&\leq \left(\sum_{j=1}^{Q(m)} \alpha_j\right) \cdot \max_{j=1,2,\ldots,Q(m)} \sum_{i\in I} y_{ji} \\
&= \max_{j=1,2,\ldots,Q(m)} \sum_{i\in I} y_{ji} = \sum_{i\in I} y_{j^*i},
\end{aligned}
$$

where the second inequality follows from taking the max of the inner sum over indices j and j^* is the index corresponding to the maximum sum.

Therefore, for any $I \subseteq \{1,2,\ldots,m\}$ with cardinality $|I| = \sqrt{m}$, there exists $j = 1,2,\ldots,Q(m)$ such that

$$
\sum_{i\in I} y_{ji} \geq m^{\epsilon}.
$$

Denote $\mathcal{F} = \left\{I \subseteq \{1,2,\ldots,m\} \mid |I| = \sqrt{m}\right\}$ which represents the set of all subsets of $\{1,2,\ldots,m\}$ with cardinality $\sqrt{m}$. Note that the cardinality of $\mathcal{F}$ is

$$
|\mathcal{F}| = \binom{m}{\sqrt{m}}.
$$

We know that for any $I \in \mathcal{F}$ there exists $\boldsymbol{y}_j \in \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots \boldsymbol{y}_{Q(m)}\}$ such that

$$
\sum_{i\in I} y_{ji} \geq m^{\epsilon}.
$$

We have $\binom{m}{\sqrt{m}}$ possibilities for I and $Q(m)$ possibilities for $\boldsymbol{y}_j$, hence by the pigeonhole principle, there exists a fixed $\boldsymbol{y} \in \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots \boldsymbol{y}_{Q(m)}\}$ and $\tilde{\mathcal{F}} \subseteq \mathcal{F}$ such that

$$
\begin{aligned}
&|\tilde{\mathcal{F}}| \geq \frac{1}{Q(m)} \binom{m}{\sqrt{m}}, \text{ and} \\
&\sum_{i\in I} y_i \geq m^{\epsilon}, \ \forall I \in \tilde{\mathcal{F}}.
\end{aligned}
\tag{4.24}
$$

Note that $\boldsymbol{y}$ is an extreme point of $\mathcal{U}$. Hence, $\boldsymbol{y}$ has exactly $\sqrt{m}$ ones and the remaining components are zeros. The maximum cardinality of subsets $I \subseteq [m]$ that can be constructed to satisfy $\sum_{i \in I} y_i \geq m^{\epsilon}$ is

$$\sum_{k=m^{\epsilon}}^{k=\sqrt{m}} \binom{\sqrt{m}}{k} \cdot \binom{m-\sqrt{m}}{\sqrt{m}-k}.$$

By over counting, the above sum can be upper-bounded by

$$\binom{\sqrt{m}}{m^{\epsilon}} \cdot \binom{m-m^{\epsilon}}{\sqrt{m}-m^{\epsilon}}.$$

Therefore, cardinality of $\tilde{\mathcal{F}}$ should be less that the above upper bound, i.e.,

$$\binom{\sqrt{m}}{m^{\epsilon}} \cdot \binom{m-m^{\epsilon}}{\sqrt{m}-m^{\epsilon}} \geq |\tilde{\mathcal{F}}| \geq \frac{1}{Q(m)} \binom{m}{\sqrt{m}}$$

Then,

$$\frac{\binom{\sqrt{m}}{m^{\epsilon}} \cdot \binom{m-m^{\epsilon}}{\sqrt{m}-m^{\epsilon}}}{\binom{m}{\sqrt{m}}} \geq \frac{1}{Q(m)}. \tag{4.25}$$

which is a contradiction. The contradiction is derived by analyzing the order of the fractions in (4.25)) (see Appendix E.9). □

4.6 Computational study

In this section, we present a computational study to compare the performance of our policy with affine policies both in terms of objective function value of problem (1.1) and computation times. We explore both cases of permutation invariant sets and non-permutations invariant sets.

4.6.1 Experimental setup

Uncertainty sets. We consider the following classes of uncertainty sets for our computational experiments.

1. **Hypersphere**. We consider the following unit hypersphere defined in (4.1),

$$\mathcal{U} = \{\boldsymbol{h} \in \mathbb{R}_+^m \mid ||\boldsymbol{h}||_2 \leq 1\}.$$

2. **p-norm balls**. We consider the following sets defined in Proposition 4.3.7.

$$\mathcal{U} = \left\{\boldsymbol{h} \in \mathbb{R}_+^m \mid ||\boldsymbol{h}||_p \leq 1\right\}.$$

   For our numerical experiments, we consider the cases of $p = 3$ and $p = 3/2$.

3. **Budget of uncertainty set**. We consider the following set defined in (4.16),

$$\mathcal{U} = \left\{\boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} h_i \leq k\right\}.$$

   Here, k denotes the budget. For our numerical experiments, we choose $k = c\sqrt{m}$ where c is a random uniform constant between 1 and 2.

4. **Intersection of budget of uncertainty sets.** We consider the following intersection of L budget of uncertainty sets:

$$\mathcal{U} = \left\{\boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{j=1}^{m} \alpha_{ij} h_j \leq 1,\ \forall i = 1,\ldots,L\right\}. \tag{4.26}$$

   Here, α_{ij} are non-negative scalars. Note that the intersection of budget of uncertainty sets are not permutation invariant. For our numerical experiments, we generate α_{ij} i.i.d. according to absolute value of standard Gaussians and we normalize $||\boldsymbol{\alpha}_i||_2$ to 1 for all i (i.e. $\boldsymbol{\alpha}_i = |\boldsymbol{G}_i|/||\boldsymbol{G}_i||_2$ where $\boldsymbol{G}_i$ are i.i.d. according to $\mathcal{N}(0, \boldsymbol{I}_m)$). We consider $L = 2$ and $L = 5$ for our experiments.

5. **Generalized budget of uncertainty set.** We consider the following set

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{\ell=1}^{m} h_\ell \leq 1 + \theta(h_i + h_j) \;\; \forall i \neq j \right\}. \tag{4.27}$$

This is a generalized version of the budget of uncertainty set (4.16) where the budget θ is not a constant but depends on the uncertain parameter $\boldsymbol{h}$. In particular, the budget in the set (4.27) depends on the sum of the two lowest components of $\boldsymbol{h}$. For our numerical experiments, we choose $\theta = O(m)$.

Instances. We construct test instances of the adjustable robust problem (1.1) as follows. We choose $n = m$, $\boldsymbol{c} = \boldsymbol{d} = \boldsymbol{e}$ and $\boldsymbol{A} = \boldsymbol{B}$ where $\boldsymbol{B}$ is randomly generated as

$$\boldsymbol{B} = \boldsymbol{I}_m + \boldsymbol{G},$$

where $\boldsymbol{I}_m$ is the identity matrix and $\boldsymbol{G}$ is a random normalized gaussian. In particular, for the hypersphere uncertainty set, the budget of uncertainty set, the intersection of budget of uncertainty sets and the generalized budget, we conisder $G_{ij} = |Y_{ij}|/\sqrt{m}$. For the 3-norm ball, $G_{ij} = |Y_{ij}|/m^{\frac{1}{3}}$ and for the $\frac{3}{2}$-norm ball, $G_{ij} = |Y_{ij}|/m^{\frac{2}{3}}$, where Y_{ij} are i.i.d. standard gaussian. We consider values of m from $m = 10$ to $m = 100$ in increments of 10 and consider 50 instances for each value of m.

Our piecewise affine policy. We construct the piecewise affine policy based on the dominating simplex $\hat{\mathcal{U}}$ as follows. For permutation invariant sets, we use the dominating simplex that can be computed in closed form. In particular, for the hypersphere uncertainty set, we use the dominating set $\hat{\mathcal{U}}$ in Proposition 4.3.6. For the p-norm balls, we use the dominating set $\hat{\mathcal{U}}$ in Proposition 4.3.7. For the budget of uncertainty set, we use the dominating set $\hat{\mathcal{U}}$ in Proposition 4.3.10 and for the generalized budget of uncertainty set (4.27), we use the dominating set $\hat{\mathcal{U}}$ in Proposition E.11.1 (see Appendix E.11).

For non-permutation invariant sets, we use Algorithm 3 to compute the dominating simplex. In particular, we get β and $\boldsymbol{v}$ that satisfies (4.5) and $2\beta \cdot \mathsf{conv}\,(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{v})$ is a dominating set (see Lemma 4.2.4-b). We can also show that the following set (4.28) is a dominating set (see Proposition E.10.1 in Appendix E.10),

$$\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\left(\boldsymbol{v}, \boldsymbol{e}_1 + \boldsymbol{v}, \ldots, \boldsymbol{e}_m + \boldsymbol{v}\right). \tag{4.28}$$

While the worst case scaling factor for the above dominating set can be 2β and therefore the theoretical bounds do not change, computationally (4.28) can provide a better policy and we use this in our numerical experiments for the intersection of budget of uncertainty sets (4.26).

4.6.2 Results

Let $z_{\mathsf{p-aff}}(\mathcal{U})$ denote the worst-case objective value of our piecewise affine police. Note that the piecewise affine policy over $\mathcal{U}$ is computed by solving the adjustable robust problem over $\hat{\mathcal{U}}$ and $z_{\mathsf{p-aff}}(\mathcal{U}) = z_{\mathsf{AR}}(\hat{\mathcal{U}})$. For each uncertainty set we report the ratio $r = \frac{z_{\mathsf{Aff}}(\mathcal{U})}{z_{\mathsf{p-aff}}(\mathcal{U})}$ for $m = 10$ to 100. In particular, for each value of m, we report the average ratio (Avg), the maximum ratio (Max), the minimum ratio (Min), the quantiles $5\%, 10\%, 25\%, 50\%$ for the ratio r, the running time of our policy ($T_{\mathsf{p-aff}}(s)$) and the running time of affine policy ($T_{\mathsf{aff}}(s)$). In addition, for the intersection of budget of uncertainty sets, we also report the computation time to construct $\hat{\mathcal{U}}$ via Algorithm 3 ($T_{\mathsf{Alg1}}(s)$). The numerical results are obtained using Gurobi 7.0.2 on a 16-core server with 2.93GHz processor and 56GB RAM.

Hypersphere and Norm-balls. We present the results of our computational experiments in Tables 4.1, 4.2 and 4.3 for the hypersphere and norm-ball uncertainty sets. We observe that the piecewise affine policy performs significantly better than affine policy for our family of test instances. In Tables 4.1, 4.2 and 4.3, we observe that the ratio $r = \frac{z_{\mathsf{Aff}}(\mathcal{U})}{z_{\mathsf{p-aff}}(\mathcal{U})}$ increases significantly as m increases which implies that our policy provides a significant improvement over affine policy for large values of m. We also observe that the ratio for the hypersphere is larger than the ratio for norm-balls. This matches the theoretical bounds presented in Table 1.2 which suggests that the improvement over affine policy is the highest for $p = 2$ for p-norm balls.

We note that for the smallest values of m ($m = 10$), the performance of affine policy is better than our policy. However, for $m > 10$, the performance of our policy is significantly better for all these three uncertainty sets: hypersphere, 3-norm ball and 3/2-norm ball.

Furthermore, our policy scales well and the average running time is less than 0.1 second even for large values of m. On the other hand, computing the optimal affine policy over $\mathcal{U}$ becomes computationally challenging as m increases. For instance, the average running time for computing an optimal affine policy for $m = 100$ is around 9 minutes for the hypersphere uncertainty set, around 17 minutes for the 3-norm ball and around 16 minutes for the $3/2$-norm ball.

Budget of uncertainty sets. We present the results of our computational experiments in Tables 4.4, 4.5, 4.6 and 4.7 for the single budget of uncertainty set, the intersection of budget sets and the generalized budget.

For the budget of uncertainty set (4.16), we observe that affine policy performs better than our piecewise affine policy for our family of test instances. Note that as we mention earlier, our policy is not a generalization of affine policies and therefore is not always better. For our experiments, we use $k = c\sqrt{m}$ which gives the worst case theoretical bound for our policy (see Theorem 4.5.1), but the performance of our policy is still reasonable and the average ratio $r = \frac{z_{\mathsf{Aff}}(\mathcal{U})}{z_{\mathsf{p-aff}}(\mathcal{U})}$ over all instances is around 0.88 as we can observe in Table 4.4. On the other hand, as in the case of conic uncertainty sets, our policy scales well with an average running time less than 0.1 second even for large values of m, whereas affine policy takes for example more than 6 minutes on average for $m = 100$.

Tables 4.5 and 4.6 present the results for intersection of budget of uncertainty sets. We observe that affine policy outperforms our policy as in the case of a single budget. This confirms that affine policy performs very well empirically for this class of uncertainty sets. We also observe that the performance of our policy improves when we increase the number of budget constraints. For example, for $m = 100$, the average ratio $r = \frac{z_{\mathsf{Aff}}(\mathcal{U})}{z_{\mathsf{p-aff}}(\mathcal{U})}$ increases from 0.79 in the case of $L = 2$ to 0.88 for $L = 5$. This suggests that the performance of our policy gets closer to the one of affine policy as long as we add more budgets constraints. While affine policy performs better than our policy for budget of uncertainty sets, we would like to note that this is not necessarily true for any polyhedral uncertainty set. In particular, we also test our policy with the generalized budget (4.27) and observe that our policy is significantly better than affine even when the set is polyhedral.

Table 4.7 presents the results for the generalized budget set (4.27). We observe that our piecewise affine policy outperforms affine policy both in terms of objective value and computation time. The gap increases as m increases which implies a significant improvement over affine policy for large values of m. Furthermore, unlike the piecewise affine policy, computing an affine solution becomes challenging for large values of m.

For the intersection of budget of uncertainty sets (4.26) that are not permutation invariant, we compute the dominating set (in particular β and $\boldsymbol{v}$) using Algorithm 3. We report the average running time, T_{Alg1} of Algorithm 3 which solves a sequence of MIPs in Tables 4.5 and 4.6. We note that there is no need to solve MIPs optimally in Algorithm 3; one can stop when a feasible solution with an objective value greater than t is found. We observe that the running time of Algorithm 3 is reasonable as compared to that of affine policy. For example, the average running time of Algorithm 3 for $m = 100$ and $L = 5$ is 7 min whereas affine policy takes 10 min in average. For large values of m and a large number of budget constraints, the running time of Algorithm 1 might increase significantly and exceed the computation time of affine policy. However, we would like to emphasize that β and $\boldsymbol{v}$ given by Algorithm 3 do not depend on the parameters $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{c}, \boldsymbol{d})$ and only depend on the uncertainty set. Therefore, they can be computed offline and can be used to solve many instances of the problem parameters for the same uncertainty set.

m	Avg	Max	Min	5%	10%	25%	50%	$T_{\text{p-aff}}(s)$	$T_{\text{aff}}(s)$
10	0.955	1.006	0.875	1.003	0.988	0.971	0.960	0.001	0.221
20	1.120	1.168	1.076	1.152	1.141	1.132	1.122	0.002	0.948
30	1.218	1.251	1.180	1.243	1.238	1.225	1.221	0.003	2.753
40	1.288	1.328	1.238	1.318	1.312	1.299	1.291	0.006	6.479
50	1.349	1.382	1.319	1.375	1.370	1.357	1.349	0.009	14.678
60	1.399	1.429	1.366	1.418	1.415	1.408	1.398	0.013	32.323
70	1.443	1.472	1.454	1.460	1.457	1.451	1.440	0.019	58.605
80	1.485	1.509	1.485	1.505	1.499	1.491	1.482	0.033	107.898
90	1.523	1.549	1.527	1.539	1.532	1.530	1.525	0.040	200.134
100	1.557	1.578	1.560	1.574	1.570	1.564	1.557	0.081	564.772

Table 4.1: Comparison on the performance and computation time of affine policy and our piecewise affine policy for the **hypersphere uncertainty set**. For 50 instances, we compute $\frac{z_{\text{Aff}}(\mathcal{U})}{z_{\text{p-aff}}(\mathcal{U})}$ and present the average, min, max ratios and the percentiles $5\%, 10\%, 25\%, 50\%$. Here, $T_{\text{p-aff}}(s)$ denotes the running time for our piecewise affine policy and $T_{\text{aff}}(s)$ denotes the running time for affine policy in seconds.

m	Avg	Max	Min	5%	10%	25%	50%	$T_{\text{p-aff}}(s)$	$T_{\text{aff}}(s)$
10	0.975	1.049	0.907	1.023	1.017	0.991	0.971	0.001	0.743
20	1.082	1.141	1.042	1.128	1.119	1.097	1.080	0.002	3.714
30	1.157	1.195	1.094	1.190	1.177	1.167	1.158	0.003	12.386
40	1.218	1.247	1.184	1.236	1.233	1.226	1.219	0.006	31.687
50	1.270	1.294	1.245	1.293	1.284	1.275	1.271	0.009	69.302
60	1.312	1.346	1.274	1.335	1.325	1.319	1.312	0.013	117.949
70	1.345	1.363	1.323	1.361	1.358	1.351	1.347	0.020	258.862
80	1.378	1.402	1.356	1.396	1.393	1.384	1.378	0.031	435.629
90	1.408	1.429	1.389	1.421	1.418	1.413	1.409	0.043	728.436
100	1.434	1.457	1.419	1.447	1.443	1.438	1.433	0.050	1033.174

Table 4.2: Comparison on the performance and computation time of affine policy and our piecewise affine policy for the 3**-norm ball uncertainty set**.

m	Avg	Max	Min	5%	10%	25%	50%	$T_{\text{p-aff}}(s)$	$T_{\text{aff}}(s)$
10	0.904	0.952	0.817	0.939	0.932	0.918	0.905	0.001	0.728
20	1.028	1.058	0.992	1.051	1.044	1.036	1.031	0.002	3.462
30	1.115	1.144	1.095	1.132	1.128	1.122	1.115	0.003	10.896
40	1.174	1.190	1.161	1.184	1.183	1.177	1.174	0.005	29.209
50	1.226	1.244	1.204	1.240	1.235	1.232	1.227	0.009	70.099
60	1.266	1.278	1.255	1.275	1.274	1.269	1.267	0.013	123.518
70	1.303	1.311	1.292	1.310	1.309	1.305	1.303	0.019	267.450
80	1.335	1.345	1.328	1.341	1.339	1.337	1.335	0.034	458.791
90	1.363	1.372	1.353	1.370	1.369	1.366	1.363	0.044	701.262
100	1.387	1.395	1.381	1.392	1.391	1.389	1.387	0.056	967.773

Table 4.3: Comparison on the performance and computation time of affine policy and our piecewise affine policy for the $3/2$**-norm ball uncertainty set.**

m	Avg	Max	Min	5%	10%	25%	50%	$T_{\text{p-aff}}(s)$	$T_{\text{aff}}(s)$
10	0.906	0.989	0.766	0.986	0.974	0.957	0.915	0.001	0.014
20	0.897	0.963	0.780	0.957	0.951	0.939	0.916	0.002	0.207
30	0.891	0.961	0.765	0.957	0.945	0.923	0.906	0.004	0.803
40	0.882	0.954	0.753	0.950	0.946	0.928	0.900	0.006	2.997
50	0.899	0.954	0.763	0.950	0.947	0.937	0.914	0.011	11.687
60	0.879	0.956	0.772	0.953	0.948	0.932	0.896	0.015	26.760
70	0.887	0.958	0.911	0.951	0.950	0.936	0.909	0.020	71.167
80	0.882	0.954	0.768	0.951	0.946	0.937	0.902	0.047	147.376
90	0.890	0.953	0.765	0.950	0.949	0.936	0.917	0.039	220.809
100	0.886	0.955	0.750	0.946	0.943	0.931	0.900	0.066	397.981

Table 4.4: Comparison on the performance and computation time of affine policy and our piecewise affine policy for the **budget of uncertainty set** with a budget $k = c\sqrt{m}$ where for each instance we generate c uniformly from $[1,2]$.

m	Avg	Max	Min	5%	10%	25%	50%	$T_{\text{p-aff}}(s)$	$T_{\text{Alg1}}(s)$	$T_{\text{aff}}(s)$
10	0.814	0.881	0.700	0.861	0.851	0.833	0.821	0.002	0.191	0.013
20	0.805	0.866	0.716	0.850	0.838	0.825	0.807	0.016	0.723	0.227
30	0.770	0.847	0.701	0.827	0.808	0.787	0.773	0.091	0.386	0.931
40	0.801	0.839	0.702	0.832	0.828	0.814	0.810	0.270	1.399	3.731
50	0.781	0.825	0.726	0.818	0.814	0.803	0.784	0.656	2.081	12.056
60	0.805	0.841	0.752	0.829	0.824	0.817	0.811	1.406	4.093	32.695
70	0.789	0.839	0.706	0.820	0.809	0.802	0.795	2.595	1.798	80.342
80	0.774	0.844	0.725	0.825	0.816	0.789	0.770	4.484	5.096	163.257
90	0.807	0.838	0.756	0.832	0.828	0.818	0.807	7.628	8.734	354.598
100	0.790	0.821	0.750	0.817	0.812	0.801	0.791	5.235	6.391	646.136

Table 4.5: Comparison on the performance and computation time of affine policy and our piecewise affine policy for the **intersection of 2 budget of uncertainty sets** (4.26).

m	Avg	Max	Min	5%	10%	25%	50%	$T_{\text{p-aff}}(s)$	$T_{\text{Alg1}}(s)$	$T_{\text{aff}}(s)$
10	0.869	0.932	0.824	0.920	0.910	0.884	0.871	0.002	0.043	0.015
20	0.852	0.924	0.795	0.909	0.893	0.870	0.852	0.021	0.058	0.309
30	0.864	0.898	0.820	0.888	0.880	0.872	0.865	0.100	0.343	1.024
40	0.856	0.896	0.802	0.883	0.882	0.874	0.861	0.290	0.464	4.010
50	0.857	0.891	0.794	0.891	0.886	0.876	0.861	0.706	3.546	12.535
60	0.880	0.900	0.860	0.894	0.892	0.885	0.881	1.471	18.474	33.693
70	0.873	0.896	0.809	0.894	0.890	0.882	0.878	2.800	13.125	82.961
80	0.858	0.889	0.825	0.886	0.881	0.872	0.858	4.809	21.780	167.753
90	0.859	0.890	0.818	0.885	0.881	0.877	0.866	8.004	144.808	344.924
100	0.885	0.902	0.865	0.900	0.896	0.893	0.888	5.821	459.436	632.483

Table 4.6: Comparison on the performance and computation time of affine policy and our piecewise affine policy for the **intersection of 5 budget of uncertainty sets** (4.26).

m	Avg	Max	Min	5%	10%	25%	50%	$T_{\text{p-aff}}(s)$	$T_{\text{aff}}(s)$
10	1.015	1.067	0.983	1.053	1.045	1.025	1.006	0.001	0.046
20	1.107	1.159	1.100	1.147	1.142	1.127	1.106	0.003	0.840
30	1.148	1.214	1.092	1.189	1.179	1.163	1.155	0.004	3.933
40	1.173	1.220	1.105	1.206	1.198	1.188	1.175	0.009	18.097
50	1.191	1.227	1.154	1.216	1.213	1.201	1.189	0.016	62.668
60	1.209	1.259	1.193	1.238	1.225	1.215	1.210	0.021	145.552
70	1.225	1.254	1.190	1.247	1.239	1.228	1.224	0.019	237.448
80	1.237	1.275	1.213	1.264	1.260	1.245	1.235	0.044	573.342
90	1.248	1.284	1.223	1.268	1.260	1.254	1.249	0.050	1168.928
100	1.257	1.274	1.240	1.271	1.268	1.261	1.257	0.053	1817.940

Table 4.7: Comparison on the performance and computation time of affine policy and our piecewise affine policy for the **generalized budget of uncertainty set** (4.27)**.**

Chapter 5: Extended affine and Threshold policies

5.1 Introduction

Recall the two-stage adjustable robust problem (1.1)[1]

$$
\begin{aligned}
z_{\mathsf{AR}}(\mathcal{U}) = \min_{\boldsymbol{x}} \ & \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \min_{\boldsymbol{y}(\boldsymbol{h})} \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{h}) \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) \ \geq \ \boldsymbol{h}, \quad \forall \boldsymbol{h} \in \mathcal{U} \\
& \boldsymbol{x} \in \mathbb{R}^n_+ \\
& \boldsymbol{y}(\boldsymbol{h}) \in \mathbb{R}^n_+, \quad \forall \boldsymbol{h} \in \mathcal{U}.
\end{aligned}
$$

In the previous chapter, we give a tractable framework to design a class of piecewise affine policies for the two-stage adjustable problem (1.1) that improves significantly over affine policy for many important uncertainty sets such as hypersphere and q-norm-balls. In this chapter, we significantly improve over the previous results and explore new approaches for designing near optimal tractable policies. In particular, we introduce *extended affine* policies and *threshold polices*. An extended affine policy is an affine policy in a lifted space, i.e., instead of restricting the second stage decision to be an affine function of the uncertain parameter $\boldsymbol{h} \in \mathcal{U}$, we first *decompose* $\mathcal{U}$ into several sets and run an affine policy in the new sets. More specifically, we present a framework where we *decompose* an uncertainty set $\mathcal{U}$ into a Minkowski sum of budget of uncertainty sets $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_L$ and define our extended affine policy as the sum of affine policies over $\mathcal{U}_j$ for $j = 1, \ldots, L$. We give an explicit construction of this decomposition for important class of uncertainty sets that can be computed efficiently. We show that our extended affine policy gives $O(\frac{\log n \log m}{\log \log n})$-approximation for the important class of permutation invariant sets that includes hypersphere and q-norm balls.

[1] Following the previous chapter, we assume that the first-stage decision $\boldsymbol{x}$ belongs to the non-negative orthant, i.e., $\boldsymbol{x} \in \mathcal{X} = \mathbb{R}^n_+$ and we refer to to $z_{\mathsf{AR}}(\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{A}, \boldsymbol{B}, \mathcal{U})$ as $z_{\mathsf{AR}}(\mathcal{U})$ for the sake of simplicity.

This approximation bound improves significantly over the previous results in the literature, for instance the best know bound in the litterature is $O\left(m^{\frac{1}{4}}\right)$ for hypersphere and $O\left(m^{\frac{q-1}{q^2}}\right)$ [41] for q-norm balls. To the best of our knowledge, the approximation bounds in this chapter are the first logarithmic approximation bounds for (1.1) under conic uncertainty sets.

In the second part of this chapter, our goal is to characterize the structure of near-optimal solutions for (1.1). In particular, we present threshold policies. These are particular class of piecewise affine policies where the second-stage decision is restricted to be of the form:

$$\boldsymbol{y}(\boldsymbol{h}) = \sum_{i=1}^{m} (h_i - \theta_i)^+ \boldsymbol{v}_i + \boldsymbol{q}.$$

Here, $\boldsymbol{\theta} \in \mathbb{R}_+^m$ is the threshold parameter, $\boldsymbol{q} \in \mathbb{R}_+^n$ and for all $i \in [m]$ $\boldsymbol{v}_i \in \mathbb{R}_+^n$. Threshold policies are widely used in practice in many settings and applications (see for instance [42]). They are highly interpretable and easy to implement in practice. However computing optimal threshold policies is often a hard problem. Based on insights from the construction of our extended affine policy, we show that the structure of a near-optimal solution for (1.1), is given by a threshold policy. In particular, we show by construction the existence of threshold policies that gives $O(\log n + \log m)$ approximation for (1.1) for hypersphere and q-norm ball uncertainty sets and give $O(\log n \log m)$-approximation for the general class of permutation invariant sets.

Following the previous chapter, we assume that the uncertainty set $\mathcal{U}$ satisfies the following assumption: $\mathcal{U} \subseteq [0,1]^m$ is convex, full-dimensional with $\boldsymbol{e}_i \in \mathcal{U}$ for all $i = 1, \ldots, m$, and *down-monotone*, i.e., $\boldsymbol{h} \in \mathcal{U}$ and $\boldsymbol{0} \leq \boldsymbol{h}' \leq \boldsymbol{h}$ implies that $\boldsymbol{h}' \in \mathcal{U}$. We would like to emphasize that the above assumption can be made without loss of generality since we can appropriately scale the uncertainty set, and consider a down-monotone completion, without affecting the two-stage problem (1.1).

The rest of this chapter is organized as follows. In Section 5.2, we present extended affine policies and show their performance for (1.1). In Section 5.3, we present our construction for threshold policies and analyze their performance for two-stage adjustable problem (1.1) under the class of permutation invariant sets.

5.2 Extended affine policies

The construction of our extended affine policy relies on our results in Chapter 3 on the performance of affine policies for budget of uncertainty sets. In fact, in Chapter 3 , we show that affine policy gives $O(\frac{\log n}{\log \log})$-approximation to (1.1) under budget of uncertainty sets which matches the hardness of approximation for (1.1) and therefore affine policy gives an optimal approximation to (1.1). The idea in this section is to decompose the uncertainty set $\mathcal{U}$ into a a Minkowski sum of small number of budget of uncertainty sets $\mathcal{U}_j$ such that each $\mathcal{U}_j$ is included in $\mathcal{U}$ and $\mathcal{U}$ is within a constant factor from $\mathcal{U}_1 \oplus \mathcal{U}_2 \ldots \oplus \mathcal{U}_L$. Our extended affine policy is defined as the sum of affine policies over the budgeted sets $\mathcal{U}_j$. More formally, let us define a *γ-budgeted decomposition* of $\mathcal{U}$ as follows.

Definition 5.2.1. Let $\mathcal{U}$ be an uncertainty set and $\gamma \geq 1$. We say that $(\mathcal{U}_1, \mathcal{U}_2, \ldots, \ldots \mathcal{U}_L)$ is a *γ-budgeted decomposition* of $\mathcal{U}$ if and only if:

- For all $j \in [L]$, $\mathcal{U}_j$ is a budget of uncertainty set.
- For all $j \in [L]$, $\mathcal{U}_j \subseteq \mathcal{U}$.
- $\mathcal{U} \subseteq \gamma \cdot \mathcal{U}_1 \oplus \mathcal{U}_2 \ldots \oplus \mathcal{U}_L$.

Recall that the definition of a budget of uncertainty set $\mathcal{V}$ is given by

$$\mathcal{V} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} w_i h_i \leq \Gamma \right\}$$

and a Minkoswki sum of sets is defined as

$$\mathcal{U}_1 \oplus \mathcal{U}_2 \ldots \oplus \mathcal{U}_L = \left\{ \sum_{j=1}^{L} \boldsymbol{h}_j \;\middle|\; \boldsymbol{h}_j \in \mathcal{U}_j,\ \forall j \in [L] \right\}.$$

Extended Affine policy. Let $\mathcal{U}$ be an uncertainty set and $(\mathcal{U}_1, \mathcal{U}_2, \ldots, \ldots \mathcal{U}_L)$ a γ-*budgeted decomposition* of $\mathcal{U}$ as defined in 5.2.1. Let $(\boldsymbol{x}^j, \boldsymbol{y}^j_{\mathsf{Aff}}(.))$ be the optimal affine policy for (1.1) under $\mathcal{U}_j$ for $j \in [L]$. Our extended affin policy is given by

$$\begin{aligned} \boldsymbol{x} &= \gamma \sum_{j=1}^{L} \boldsymbol{x}^j \\ \boldsymbol{y}(\boldsymbol{h}) &= \gamma \sum_{j=1}^{L} \boldsymbol{y}^j_{\mathsf{Aff}}(\boldsymbol{h}_j) \end{aligned} \tag{5.1}$$

where $\boldsymbol{h} = \gamma \sum_{j=1}^{L} \boldsymbol{h}_j$ for some $\boldsymbol{h}_j \in \mathcal{U}_j$, $j \in [L]$.

It is clear that the extended affine policy defined in (5.1) is feasible for (1.1). Moreover, it can be computed efficiently by solving the affine policies over the sets $\mathcal{U}_j$, $j \in [L]$. Note that an affine policy over $\mathcal{U}_j$ can be computed by solving a single compact LP (see Chapter 1). In the following theorem, we show that our extended policy gives $O\left(\frac{\gamma L \log n}{\log\log n}\right)$-approximation to (1.1).

Theorem 5.2.2. *Let $\mathcal{U}$ be an uncertainty set and $(\mathcal{U}_1, \mathcal{U}_2, \ldots, \ldots \mathcal{U}_L)$ a γ-budgeted decomposition of $\mathcal{U}$ as defined in 5.2.1. The extended affine policy defined in* (5.1) *gives $O\left(\frac{\gamma L \log n}{\log\log n}\right)$-approximation to* (1.1)*, i.e.,*

$$\frac{1}{\gamma} z_{\mathsf{AR}}(\mathcal{U}) \leq \sum_{j=1}^{L} z_{\mathsf{Aff}}(\mathcal{U}_j) \leq O\left(\frac{L \log n}{\log\log n}\right) \cdot z_{\mathsf{AR}}(\mathcal{U}),$$

where $z_{\mathsf{Aff}}(\mathcal{U}_j)$ is the cost of the optimal affine policy over $\mathcal{U}_j$.

Proof. Let $\boldsymbol{h} \in \mathcal{U}$. Then, there exists $\boldsymbol{h}_j \in \mathcal{U}_j$ for $j \in [L]$, such that $\boldsymbol{h} = \gamma \sum_{j=1}^{L} \boldsymbol{h}_j$. For $j \in [L]$, consider $(\boldsymbol{x}^*_j, \boldsymbol{y}^*_j(.))$ an optimal solution for $z_{\mathsf{AR}}(\mathcal{U}_j)$. Therefore,

$$\gamma \sum_{j=1}^{L} \left(\boldsymbol{A}\boldsymbol{x}^*_j + \boldsymbol{B}\boldsymbol{y}_j(\boldsymbol{h}_j) \right) \geq \gamma \sum_{j=1}^{L} \boldsymbol{h}_j = \boldsymbol{h}.$$

Hence, $(\gamma \sum_{j=1}^{L} \boldsymbol{x}_j^*, \gamma \sum_{j=1}^{L} \boldsymbol{y}_j^*(.))$ is a feasible solution for $z_{\mathsf{AR}}(\mathcal{U})$ and therefore

$$z_{\mathsf{AR}}(\mathcal{U}) \leq \gamma \sum_{j=1}^{L} z_{\mathsf{AR}}(\mathcal{U}_j).$$

Moreover, we know that $z_{\mathsf{AR}}(\mathcal{U}_j) \leq z_{\mathsf{Aff}}(\mathcal{U}_j)$ which implies the first inequality. On the other hand, we know from the main result in Chapter 3 (Theorem 3.2.1) that affine policy gives $O(\frac{\log n}{\log \log n})$-approximation for (1.1) under budget of uncertainty sets. Therefore,

$$z_{\mathsf{Aff}}(\mathcal{U}_j) \leq O\left(\frac{\log n}{\log \log n}\right) \cdot z_{\mathsf{AR}}(\mathcal{U}_j).$$

Moreover, since $\mathcal{U}_j \subseteq \mathcal{U}$, then $z_{\mathsf{AR}}(\mathcal{U}_j) \leq z_{\mathsf{AR}}(\mathcal{U})$. By taking the sum over j, we get the second inequality. $\square$

In this chapter, we focus on the class of permutation invariant sets. This is a class that includes many important sets used in the literature of robust optimization.

Definition 5.2.3 (Permutation invariant sets)**.** $\mathcal{U}$ is said to be permutation invariant if $\boldsymbol{h} \in \mathcal{U}$ implies that for any permutation σ of $\{1, ..., m\}$, $\boldsymbol{h}^{\sigma} \in \mathcal{U}$ where for all $i \in [m]$ $h_i^{\sigma} = h_{\sigma(i)}$.

This class of uncertainty sets contains in particular q-norm balls. Note that the best known bound for approximating (1.1) under q-norm balls is $O(m^{\frac{q-1}{q^2}})$ [41] and the best known bound for approximating (1.1) under q-norm balls using affine policies is $O(m^{\frac{1}{q}})$ [36].

We present an explicit construction of an extended affine policy that gives $O(\frac{\log n \log m}{\log \log n})$ approximation for (1.1) for permutation invariant sets. This improves scientifically over the previous known bound and almost matches the hardness of the problem. It is sufficient to show the existence of a γ-budgeted decomposition where γ is a constant and the number of budgeted sets in the decomposition is $O(\log m)$ and then apply Theorem 5.2.2. In particular, let $\mathcal{U}$ be an uncertainty set that is permutation invariant. For all $j \in \{0, \ldots, \lceil \log m \rceil\}$, define the following budget of uncertainty sets

$$\hat{\mathcal{U}}_j = \left\{ \boldsymbol{h} \in \left[0, \frac{1}{2^j}\right]^m \,\middle|\, \sum_{i=1}^{m} h_i \leq \frac{k_j}{2^j} \right\}, \tag{5.2}$$

where

$$k_j = \max \left\{ k \in [m] \,\middle|\, \frac{1}{2^j} \sum_{i=1}^{k} \boldsymbol{e}_i \in \mathcal{U} \right\}. \tag{5.3}$$

Intuitively, k_j can be seen as how many components in $\mathcal{U}$ can be equal to $\frac{1}{2^j}$.

Claim 5.2.4. *For all* $0 \leq j \leq \lceil \log m \rceil$*, we have* $\hat{\mathcal{U}}_j \subset \mathcal{U}$.

Proof. Let $0 \leq j \leq \lceil \log m \rceil$. It suffices to show that all extreme points of $\hat{\mathcal{U}}_j$ are in $\mathcal{U}$. Let us rewrite the set $\hat{\mathcal{U}}_j$ as follows

$$\hat{\mathcal{U}}_j = \frac{1}{2^j} \mathcal{V}_j$$

where

$$\mathcal{V}_j = \left\{ \boldsymbol{h} \in [0,1]^m \,\middle|\, \sum_{i=1}^{m} h_i \leq k_j \right\}.$$

Since k_j is an integer, $\mathcal{V}_j$ is a k-ones set and all its extreme points are in $\{0,1\}^m$. Therefore, all the pareto extreme points of $\hat{\mathcal{U}}_j$ are of the form $\boldsymbol{s}_j^{\sigma} = \frac{1}{2^j} \sum_{i=1}^{k_j} \boldsymbol{e}_{\sigma(i)}$ where $\sigma \in S^m$ and S^m is the set of permutations of $[m]$. Note that by definition of k_j, there exists a permutation $\sigma_0 \in S^m$ such that $\boldsymbol{s}_j^{\sigma_0} \in \mathcal{U}$. Since $\mathcal{U}$ is permutation invariant, this implies that $\boldsymbol{s}_j^{\sigma} \in \mathcal{U}$ for all $\sigma \in S^m$. Hence $\mathcal{U}$ contains all pareto extreme points of $\hat{\mathcal{U}}_j$ and therefore by down-monotonicity $\hat{\mathcal{U}}_j \subset \mathcal{U}$. □

Claim 5.2.5. *We have* $\mathcal{U} \subseteq 2 \cdot \hat{\mathcal{U}}_0 \oplus \hat{\mathcal{U}}_2 \ldots \oplus \hat{\mathcal{U}}_{\lceil \log m \rceil}$.

Proof. Let $\boldsymbol{h} \in \mathcal{U}$ and $L = \lceil \log m \rceil$. We have

$$\begin{aligned}\boldsymbol{h} &= \sum_{i=1}^{m} h_i \boldsymbol{e}_i \\ &= \sum_{i=1}^{m} \left(\sum_{j=0}^{L-1} \mathbf{1}\left(\frac{1}{2^j} \le h_i < \frac{1}{2^{j-1}} \right) + \mathbf{1}\left(h_i < \frac{1}{2^{L-1}} \right) \right) h_i \boldsymbol{e}_i \\ &= \sum_{j=0}^{L} \hat{\boldsymbol{h}}_j,\end{aligned}$$

where for $0 \le j \le L-1$,

$$\hat{\boldsymbol{h}}_j = \sum_{i=1}^{m} \mathbf{1}\left(\frac{1}{2^j} \le h_i < \frac{1}{2^{j-1}} \right) h_i \boldsymbol{e}_i$$

and

$$\hat{\boldsymbol{h}}_L = \sum_{i=1}^{m} \mathbf{1}\left(h_i < \frac{1}{2^{L-1}} \right) h_i \boldsymbol{e}_i.$$

We have for $0 \le j \le L-1$,

$$\hat{\boldsymbol{h}}_j \ge \frac{1}{2^j} \cdot \sum_{i=1}^{m} \mathbf{1}\left(\frac{1}{2^j} \le h_i < \frac{1}{2^{j-1}} \right) \boldsymbol{e}_i.$$

By down-monotonicity of $\mathcal{U}$, we get

$$\frac{1}{2^j} \cdot \sum_{i=1}^{m} \mathbf{1}\left(\frac{1}{2^j} \le h_i < \frac{1}{2^{j-1}} \right) \boldsymbol{e}_i \in \mathcal{U}$$

and therefore

$$\sum_{i=1}^{m} \mathbf{1}\left(\frac{1}{2^j} \le h_i < \frac{1}{2^{j-1}} \right) \le k_j.$$

Hence,

$$\sum_{i=1}^{m} \mathbf{1}\left(\frac{1}{2^j} \le h_i < \frac{1}{2^{j-1}} \right) h_i \le \frac{k_j}{2^{j-1}},$$

which implies that $\hat{\boldsymbol{h}}_j \in 2 \cdot \hat{\mathcal{U}}_j$ for all $0 \le j \le L-1$. Moreover,

$$\hat{\boldsymbol{h}}_L \le \sum_{i=1}^{m} \mathbf{1}\left(h_i < \frac{1}{2^{L-1}}\right) \frac{1}{2^{L-1}} \boldsymbol{e}_i \le \sum_{i=1}^{m} \mathbf{1}\left(h_i < \frac{1}{2^{L-1}}\right) \frac{2}{m} \boldsymbol{e}_i \le \frac{2}{m} \boldsymbol{e} \in 2\hat{\mathcal{U}}_L.$$

Therefore, $\hat{\boldsymbol{h}}_L \in 2\hat{\mathcal{U}}_L$ and

$$\boldsymbol{h} \in 2 \cdot \hat{\mathcal{U}}_0 \oplus \hat{\mathcal{U}}_2 \ldots \oplus \hat{\mathcal{U}}_L.$$

□

From Claim 5.2.4 and Claim 5.2.5, we conclude that $(\hat{\mathcal{U}}_0, \hat{\mathcal{U}}_2 \ldots, \hat{\mathcal{U}}_{\lceil \log m \rceil})$ is a 2-budgeted decomposition of $\mathcal{U}$ and therefore by applying the result in Theorem 5.2.2, we get $O(\frac{\log n \log m}{\log \log n})$-approximation to (1.1). In particular, we have the following theorem.

Theorem 5.2.6. *Let $\mathcal{U}$ be an uncertainty set that is permutation invariant. Then, our extended affine policy* (5.1) *with the budgeted decomposition defined in* (5.2) *gives $O(\frac{\log n \log m}{\log \log n})$-approximation for the adjustable problem* (1.1).

5.3 Threshold policies

In this section, we introduce threshold policies and study their properties and performance for the two-stage adjustable problem (1.1). Let $\mathcal{U}$ be an uncertainty set. Threshold policies are piecewise affine policies of the form:

$$\boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{P}(\boldsymbol{h} - \boldsymbol{\theta})^+ + \boldsymbol{q} \qquad \forall \boldsymbol{h} \in \mathcal{U},$$

where $\boldsymbol{\theta} \in \mathbb{R}^m_+$, $\boldsymbol{P} \in \mathbb{R}^{n \times m}_+$ and $\boldsymbol{q} \in \mathbb{R}^n_+$. Note that $(\boldsymbol{h} - \boldsymbol{\theta})^+$ is a vector in $\mathbb{R}^m$ of i-th coordinate $(h_i - \theta_i)^+$. Threshold policies are a generalization of affine policies, where the recourse decision is a threshold function. In particular, for each component, the policy is affine in the i-th coordinate h_i of the uncertain parameter if h_i exceeds a threshold θ_i and static otherwise. In our problem (1.1), threshold policies with a threshold parameter θ consists of covering the threshold using a static solution and the residual demand $(\boldsymbol{h} - \boldsymbol{\theta})^+$ using an affine solution. However, it is hard in general to compute optimal threshold policies due to the non linearity in $(\boldsymbol{h} - \boldsymbol{\theta})^+$. Here, our goal is not to compute optimal threshold policies, but to analyze the structure of a near-optimal policy for (1.1) and show that it could be captured by a threshold policy. In particular, we show the existence of threshold policies that give a logarithmic approximation bound to the two-stage adjustable problem (1.1) and almost matches the hardness of (1.1) under the important class of permutation invariant sets. We present an explicit construction of threshold policies based on our insights from the previous section on extended affine policies. However, the construction needs to guess the value or an approximate value of OPT.

Consider the two-stage adjustable robust optimization problem with covering constraints (1.1). When we restrict the second-stage decision to be a threshold policy, the problem becomes,

$$
\begin{aligned}
z_{\mathsf{T}}(\mathcal{U}) = \min_{\boldsymbol{x}} \ & \boldsymbol{c}^T \boldsymbol{x} + \max_{\boldsymbol{h} \in \mathcal{U}} \min_{\boldsymbol{y}(\boldsymbol{h})} \boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{h}) \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) \ \geq \ \boldsymbol{h}, \quad \forall \boldsymbol{h} \in \mathcal{U} \\
& \boldsymbol{y}(\boldsymbol{h}) = \sum_{i=1}^{m} (h_i - \theta_i)^+ \boldsymbol{p}_i + \boldsymbol{q}, \ \ \forall \boldsymbol{h} \in \mathcal{U} \\
& \boldsymbol{x} \in \mathbb{R}^n_+, \ \boldsymbol{\theta} \in \mathbb{R}^m_+, \boldsymbol{q} \in \mathbb{R}^n_+, \boldsymbol{p}_i \in \mathbb{R}^n_+, \ \forall i \in [m].
\end{aligned}
\tag{5.4}
$$

Again, we focus on the class of permutation invariant sets. In the section, we show that threshold policies provide strong performance bounds that are logarithmic in the dimension of the problem for the class of permutation invariant sets. The performance of the threshold policy depends on a factor $\tau(U)$ that characterizes the geometry of the set $\mathcal{U}$. This factor ranges between 1 and $O(\log m)$. We present the exact definition of $\tau(U)$ later on. The main result in this section is presented in the following theorem.

Theorem 5.3.1. *Consider the two-stage adjustable problem* (1.1) *where* $\mathcal{U}$ *is permutation invariant set. Then,*

$$z_{\mathsf{T}}(\mathcal{U}) = O\left(\tau(\mathcal{U}) \log n + \log m\right) z_{\mathsf{AR}}(\mathcal{U})$$

where $\tau(\mathcal{U}) = O(\log m)$ is a factor that depends on the geometry of the set $\mathcal{U}$.

For q-norm balls, we show that $\tau(\mathcal{U}) \leq 2$, therefore we have the following corollary.

Corollary 5.3.2. *Consider the two-stage adjustable problem* (1.1) *under q-norm ball uncertainty set, i.e., $\mathcal{U} = \{\boldsymbol{h} \in \mathbb{R}^m_+ \mid \sum_{i=1}^m h_i^q \leq 1\}$ where $q \geq 1$. Then,*

$$z_{\mathsf{T}}(\mathcal{U}) = O\left(\log n + \log m\right) z_{\mathsf{AR}}(\mathcal{U}).$$

Theorem 5.3.1 implies that threshold policies are at most within $O(\log m \log n)$ from the optimal solution to (1.1) in the general case of permutation invariant sets. We prove Theorem 5.3.1 by constructing explicitly the threshold policy giving a guess on the value of OPT.

5.3.1 Construction of the threshold policy

In a first part, we present the global structure of our threshold policy and show its feasibility. We then specify how to explicitly construct the threshold parameter $\boldsymbol{\theta}$ given an uncertainty set $\mathcal{U}$ and show that our choice lead to a near-optimal threshold policy. Let $\boldsymbol{\theta} \in \mathbb{R}^m_+$ and for all $i \in [m]$, define $\boldsymbol{v}_i$ as the optimal static decision to cover requirement $\boldsymbol{e}_i$, i.e.,

$$\boldsymbol{v}_i \in \operatorname{argmin}_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{e}_i \right\}. \tag{5.5}$$

Our policy is composed of a piecewise linear part and a static part.

Piecewise Linear Part.

$$\boldsymbol{y}_{\mathsf{PL}}(\boldsymbol{h}) = \sum_{i=1}^{m} (h_i - \theta_i)^+ \boldsymbol{v}_i, \qquad \forall \boldsymbol{h} \in \mathcal{U}. \tag{5.6}$$

Static Part.

$$(\boldsymbol{x}_{\mathsf{Sta}}, \boldsymbol{y}_{\mathsf{Sta}}) \in \operatorname{argmin}_{\boldsymbol{x},\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \sum_{i=1}^{m} \theta_i \boldsymbol{e}_i \right\}. \tag{5.7}$$

Threshold Policy. Our threshold policy is given by,

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{x}_{\mathsf{Sta}} \\ \boldsymbol{y}(\boldsymbol{h}) &= \boldsymbol{y}_{\mathsf{PL}}(\boldsymbol{h}) + \boldsymbol{y}_{\mathsf{Sta}}, \ \ \forall \boldsymbol{h} \in \mathcal{U}. \end{aligned} \tag{5.8}$$

Lemma 5.3.3. *The threshold policy given in* (5.8) *is feasible for problem* (1.1).

Proof. Let $\boldsymbol{h}$ be in $\mathcal{U}$. For all $i \in [m]$ we have,

$$\begin{aligned} (\boldsymbol{Ax} + \boldsymbol{By}(\boldsymbol{h}))_i &= (\boldsymbol{Ax}_{\mathsf{Sta}} + \boldsymbol{By}_{\mathsf{Sta}})_i + (\boldsymbol{By}_{\mathsf{PL}}(\boldsymbol{h}))_i \\ &\geq \theta_i + (h_i - \theta_i)^+ \\ &\geq h_i. \end{aligned}$$

Therefore, our policy is feasible. □

Construction of threshold $\boldsymbol{\theta}$.

The threshold policy in (5.8) is feasible for any value of $\boldsymbol{\theta}$. Here, we present the construction of the parameter $\boldsymbol{\theta}$ for which the threshold policy gives $O\left(\tau(\mathcal{U})\log n + \log m\right)$-approximation to (1.1) and therefore show Theorem 5.3.1. Consider an uncertainty set $\mathcal{U}$ that is permutation invariant. Recall the coefficient k_j defined in (5.3), for all $j \in \{0, \ldots, \lceil \log m \rceil\}$,

$$k_j = \max\left\{ k \in [m] \;\middle|\; \frac{1}{2^j} \sum_{i=1}^{k} \boldsymbol{e}_i \in \mathcal{U} \right\}.$$

Intuitively, k_j can be seen as how many components in $\mathcal{U}$ can be equal to $\frac{1}{2^j}$. Define

$$j_{\mathsf{max}} = \min\{ j \geq 0 \mid k_j = m \}.$$

Intuitively, j_{max} is the smallest j such that the hypercube $\left[0, \frac{1}{2^j}\right]^m \subset \mathcal{U}$. By assumption, $\boldsymbol{e}_i \in \mathcal{U}$ for all $i \in [m]$, hence by convexity and down-monotonicity, $\left[0, \frac{1}{m}\right]^m \subset \mathcal{U}$ and therefore j_{max} is well defined and $j_{\mathsf{max}} \leq \lceil \log m \rceil$.

Lemma 5.3.4. *Let k_j be as defined in* (5.3). *Then, the sequence $\left(\frac{2^j}{k_j}\right)_{0 \leq j < j_{\mathsf{max}}}$ is decreasing.*

Proof. The lemma is equivalent to show that $k_{j+1} \geq 2k_j$ for all j such that $j + 1 < j_{\mathsf{max}}$. By definition of k_j, we have,

$$\sum_{\ell=1}^{k_j} \frac{1}{2^j} \boldsymbol{e}_\ell \in \mathcal{U}.$$

Because $\mathcal{U}$ is permutation invariant and down-monotone, we get

$$\sum_{\ell=k_j+1}^{\min(2k_j,m)} \frac{1}{2^j} \boldsymbol{e}_\ell \in \mathcal{U}.$$

Therefore by convexity of $\mathcal{U}$, we have

$$\frac{1}{2} \sum_{\ell=1}^{k_j} \frac{1}{2^j} \boldsymbol{e}_\ell + \frac{1}{2} \sum_{\ell=k_j+1}^{\min(2k_j,m)} \frac{1}{2^j} \boldsymbol{e}_\ell = \sum_{\ell=1}^{\min(2k_j,m)} \frac{1}{2^{j+1}} \boldsymbol{e}_\ell \in \mathcal{U}.$$

This implies by definition of k_{j+1} that

$$k_{j+1} \geq \min(2k_j, m).$$

If $m < 2k_j$, then $k_{j+1} \geq m$, which contradicts the fact that $j + 1 < j_{\mathsf{max}}$. Hence, $m \geq 2k_j$ and therefore $k_{j+1} \geq 2k_j$.

□

From now on, we set

$$(\alpha, \beta) = (8, 8 \log n).$$

We consider an optimal solution $\boldsymbol{x}^*, \boldsymbol{y}^*(\boldsymbol{h})$ where $\boldsymbol{h} \in \mathcal{U}$, for the adjustable problem (1.1). Let OPT be the optimal cost for (1.1) and $\text{OPT}_1, \text{OPT}_2$ respectively the first stage cost and the second-stage cost associated with $\boldsymbol{x}^*, \boldsymbol{y}^*(\boldsymbol{h})$, i.e.,

$$\begin{aligned}
\text{OPT}_1 &= \boldsymbol{c}^T \boldsymbol{x}^* \\
\text{OPT}_2 &= \max_{\boldsymbol{h}\in\mathcal{U}} \ \boldsymbol{d}^T \boldsymbol{y}^*(\boldsymbol{h}) \\
\text{OPT} &= \text{OPT}_1 + \text{OPT}_2 = z_{\mathsf{AR}}(\mathcal{U}).
\end{aligned}$$

For all $i \in [m]$ define,

$$J_i = \max\left\{0 \le j \le j_{\mathsf{max}} \;\middle|\; \boldsymbol{d}^T \boldsymbol{v}_i \le \frac{2^j}{k_j}\beta \cdot \text{OPT}\right\}$$

where $\boldsymbol{v}_i$ is defined in (5.5) and the convention $\max \emptyset = -\infty$. Now set for all $i \in [m]$,

$$\theta_i = \begin{cases} 1 & J_i = -\infty \\ \frac{1}{2^{J_i}} & 0 \le J_i \le j_{\mathsf{max}} \end{cases} \tag{5.9}$$

5.3.2 Cost analysis

Let us analyze the cost of the threshold policy in (5.8) where the threshold parameter $\boldsymbol{\theta}$ is given by (5.9). We first analyze the cost of the piecewise linear part (5.6) and then the cost of the static part (5.7). But first, let us introduce the following notations and definitions.

We denote $z(\boldsymbol{h})$, the cost of covering the requirement $\boldsymbol{h}$ in the second-stage, i.e.,

$$z(\boldsymbol{h}) = \min_{\boldsymbol{y}\ge\boldsymbol{0}} \left\{\boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \boldsymbol{B}\boldsymbol{y} \ge \boldsymbol{h}\right\}. \tag{5.10}$$

Our cost analysis depends on a constant $\tau(U)$ that characterizes the geometry of the uncertainty set $\mathcal{U}$ which is defined as follows.

Definition 5.3.5. The geometric factor $\tau(\mathcal{U})$ is defined as,

$$\tau(\mathcal{U}) = \max_{\boldsymbol{h}\in\mathcal{U}} \quad \sum_{j=0}^{j_{\mathsf{max}}} \frac{u_j(\boldsymbol{h})}{k_j} \tag{5.11}$$

where for $0 \leq j \leq j_{\mathsf{max}} - 1$,

$$u_j(\boldsymbol{h}) = \sum_{i=1}^{m} \mathbf{1}\left(\frac{1}{2^j} \leq h_i < \frac{1}{2^{j-1}}\right)$$

and

$$u_{j_{\mathsf{max}}}(\boldsymbol{h}) = \sum_{i=1}^{m} \mathbf{1}\left(h_i < \frac{1}{2^{j_{\mathsf{max}}-1}}\right).$$

Note that, by definition, k_j is the maximum number of $\frac{1}{2^j}$ that we can fit in $\mathcal{U}$. Hence, $u_j(\boldsymbol{h}) \leq k_j$ and therefore $\tau(\mathcal{U}) \leq j_{\mathsf{max}} + 1 \leq \lceil \log m \rceil + 1$. In particular, we have the following lemma.

Lemma 5.3.6. *For any uncertainty set $\mathcal{U}$ that is permutation invariant, we have $\tau(\mathcal{U}) = O(\log m)$.*

Example. For q-norm balls, i.e., $\mathcal{U} = \{\boldsymbol{h} \in \mathbb{R}_+^m \mid \sum_{i=1}^m h_i^q \leq 1\}$ where $q \geq 1$, we have $\tau(\mathcal{U}) \leq 2$. In fact, for q-norm balls we have, for all j, $k_j = 2^{jq}$. Therefore, for any $\boldsymbol{h} \in \mathcal{U}$,

$$\sum_{j=0}^{j_{\mathsf{max}}} \frac{u_j(\boldsymbol{h})}{(2^j)^q} = \sum_{i=1}^{m} \sum_{j=0}^{j_{\mathsf{max}}-1} \mathbf{1}\left(\frac{1}{2^j} \leq h_i < \frac{1}{2^{j-1}}\right)\frac{1}{2^{jq}} + \mathbf{1}\left(h_i < \frac{1}{2^{j_{\mathsf{max}}-1}}\right)\frac{1}{m} \leq \sum_{i=1}^{m} h_i^q + \sum_{i=1}^{m} \frac{1}{m} \leq 2.$$

Cost of the piecewise linear part. The cost of the piecewise linear part depends on the geometric factor $\tau(\mathcal{U})$. In particular, we have the following lemma.

Lemma 5.3.7. *The cost of the piecewise linear part* (5.6) *is bounded by* $2\tau(\mathcal{U}) \cdot \beta \cdot \mathrm{OPT}$.

Proof.

$$
\begin{aligned}
\boldsymbol{d}^T y_{\mathsf{PL}}(\boldsymbol{h}) &= \sum_{i=1}^{m} (h_i - \theta_i)^+ \boldsymbol{d}^T \boldsymbol{v}_i \\
&= \sum_{i \in [m], J_i \geq 0} (h_i - \theta_i)^+ \boldsymbol{d}^T \boldsymbol{v}_i \\
&\leq \sum_{i \in [m], J_i \geq 0} \left(h_i - \frac{1}{2^{J_i}} \right)^+ \cdot \frac{2^{J_i}}{k_{J_i}} \cdot \beta \mathrm{OPT} \\
&= \sum_{i \in [m], J_i \geq 0} \sum_{j=0}^{j_{\mathsf{max}}} \left(\mathbf{1}\left(\frac{1}{2^j} \leq h_i < \frac{1}{2^{j-1}} \right) + \mathbf{1}\left(h_i < \frac{1}{2^{j_{\mathsf{max}}-1}} \right) \right) \left(h_i - \frac{1}{2^{J_i}} \right)^+ \cdot \frac{2^{J_i}}{k_{J_i}} \cdot \beta \mathrm{OPT}
\end{aligned}
$$

where the first equality holds because if $J_i = -\infty$, then $\theta_i = 1$ and therefore $(h_i - \theta_i)^+ = 0$. The first inequality follows from the definition of J_i. We have for $0 \leq j \leq j_{\mathsf{max}} - 1$,

$$
\begin{aligned}
\mathbf{1}\left(\frac{1}{2^j} \leq h_i < \frac{1}{2^{j-1}} \right) \left(h_i - \frac{1}{2^{J_i}} \right)^+ \cdot \frac{2^{J_i}}{k_{J_i}} &\leq \mathbf{1}\left(\frac{1}{2^j} \leq h_i < \frac{1}{2^{j-1}} \right) \frac{1}{2^{j-1}} \mathbf{1}(j \leq J_i) \frac{2^{J_i}}{k_{J_i}} \\
&\leq \mathbf{1}\left(\frac{1}{2^j} \leq h_i < \frac{1}{2^{j-1}} \right) \frac{1}{2^{j-1}} \frac{2^j}{k_j} \\
&= \mathbf{1}\left(\frac{1}{2^j} \leq h_i < \frac{1}{2^{j-1}} \right) \frac{2}{k_j}
\end{aligned}
$$

where the second inequality follows from Lemma 5.3.4. Moreover, for $j = j_{\mathsf{max}}$,

$$
\begin{aligned}
\mathbf{1}\left(h_i < \frac{1}{2^{j_{\mathsf{max}}-1}} \right) \left(h_i - \frac{1}{2^{J_i}} \right)^+ \cdot \frac{2^{J_i}}{k_{J_i}} &\leq \mathbf{1}\left(h_i < \frac{1}{2^{j_{\mathsf{max}}-1}} \right) \frac{1}{2^{j_{\mathsf{max}}-1}} \mathbf{1}(j_{\mathsf{max}} \leq J_i) \frac{2^{J_i}}{k_{J_i}} \\
&\leq \mathbf{1}\left(h_i < \frac{1}{2^{j_{\mathsf{max}}-1}} \right) \frac{1}{2^{j_{\mathsf{max}}-1}} \frac{2^{j_{\mathsf{max}}}}{k_{j_{\mathsf{max}}}} \\
&= \mathbf{1}\left(h_i < \frac{1}{2^{j_{\mathsf{max}}-1}} \right) \frac{2}{k_{j_{\mathsf{max}}}}.
\end{aligned}
$$

Therefore,

$$
\boldsymbol{d}^T y_{\mathsf{PL}}(\boldsymbol{h}) \leq \sum_{i \in [m]} \sum_{j=0}^{j_{\mathsf{max}}} \left(\mathbf{1}\left(\frac{1}{2^j} \leq h_i < \frac{1}{2^{j-1}} \right) \frac{2}{k_j} + \mathbf{1}\left(h_i < \frac{1}{2^{j_{\mathsf{max}}-1}} \right) \frac{2}{k_{j_{\mathsf{max}}}} \right) \cdot \beta \mathrm{OPT} \leq 2\tau(\mathcal{U}) \cdot \beta \mathrm{OPT},
$$

where the last inequality follows exactly from the definition of $\tau(\mathcal{U})$.

$\square$

Cost of static part.

Lemma 5.3.8. *The cost of the static part* (5.7) *is bounded by* $O(\alpha \log m) \cdot \mathrm{OPT}$.

Before showing Lemma 5.3.8, let us introduce the following definitions and useful lemmas that we would need for the proof. For $0 \leq j \leq j_{\mathsf{max}}$, define

$$\mathcal{I}_j = \left\{ i \in [m] \;\middle|\; \theta_i = \frac{1}{2^j} \right\},$$

$$\mathcal{J}_{1j} = \left\{ i \in [m] \;\middle|\; (\boldsymbol{A}\boldsymbol{x}^*)_i \geq \frac{1}{4} \cdot \frac{1}{2^j} \right\},$$

$$\mathcal{J}_{2j} = \mathcal{I}_j \setminus \mathcal{J}_{1j}.$$

and for $0 \leq j \leq j_{\mathsf{max}}$, define the following budget of uncertainty sets

$$\hat{\mathcal{U}}_j = \left\{ \boldsymbol{h} \in \left[0, \frac{1}{2^j}\right]^m \;\middle|\; \sum_{i=1}^{m} h_i \leq \frac{k_j}{2^j} \right\}. \tag{5.12}$$

We know from Claim 5.2.4 that for all $0 \leq j \leq j_{\mathsf{max}}$, we have $\hat{\mathcal{U}}_j \subset \mathcal{U}$.

Our proof relies on the structural result, Lemma 3.3 in El Housni and Goyal [54] which we have shown in Chapter 2 of this thesis. We restate the lemma here in our context for completeness.

Lemma 5.3.9 (Lemma 3.3 in [54])**.** *Consider* $\boldsymbol{B} \in \mathbb{R}_+^{m \times n}$, $\boldsymbol{d} \in \mathbb{R}_+^n$ *and* $\mathcal{J} \subseteq [m]$. *Let* $z(\boldsymbol{h})$ *be the cost of covering* $\boldsymbol{h}$ *as defined in* (5.10). *Suppose there exists* $\gamma > 0$ *and* $0 \leq k \leq m$ *such that the following two conditions are satisfied:*

1. *for all* $i \in \mathcal{J}$,
$$z(\boldsymbol{e}_i) > \frac{\gamma}{k} \cdot \frac{\beta}{2},$$

2. *for all* $\mathcal{W} \subseteq \mathcal{J}$,
$$|\mathcal{W}| \leq k \;\; \textit{implies} \;\; z(\mathcal{W}) \leq \gamma.$$

Then,

$$z(\mathcal{J}) \le \gamma \cdot \frac{\alpha}{2}.$$

Note that in [54], the above lemma has been shown for $\alpha = \beta = \frac{8\log n}{\log\log n}$, the lemma is also correct for $(\alpha, \beta) = (8, 8\log n)$ and the proof is along the same lines as in [54]. Now, we are ready to show Lemma 5.3.8.

Proof of Lemma 5.3.8. Fix j such that $0 \le j \le j_{\max}$. We have for all $i \in \mathcal{J}_{1j}$, $4\boldsymbol{A}\boldsymbol{x}^* \ge \frac{1}{2^j}\boldsymbol{e}_i$ and we know that $4\boldsymbol{c}^T\boldsymbol{x}^* = 4\mathrm{OPT}_1$. Therefore we can cover the components of $\mathcal{J}_{1j}$ using the static solution $4\boldsymbol{x}^*$ and pay a cost $4\mathrm{OPT}_1$. We apply Lemma 5.3.9 with $\mathcal{J}_{2j}$ and $\gamma = 2^j \cdot 4\mathrm{OPT}$. In fact, because $\mathcal{J}_{2j} \subseteq \mathcal{I}_j$, we have for all $i \in \mathcal{J}_{2j}$, $\theta_i = \frac{1}{2^j}$, which implies

$$z(\boldsymbol{e}_i) = \boldsymbol{d}^T\boldsymbol{v}_i > \frac{2^{j+1}}{k_{j+1}}\beta \cdot \mathrm{OPT} = \frac{\gamma \cdot \beta}{2k_{j+1}}$$

i.e., the first condition of Lemma 5.3.9 is satisfied with $k = k_{j+1}$.

Moreover, let $\mathcal{W} \subseteq \mathcal{J}_{2j}$ such that $|\mathcal{W}| \le k_{j+1}$. Denote by $\mathbb{1}(\mathcal{W})$ the sum $\sum_{i\in\mathcal{W}} \boldsymbol{e}_i$ and $\boldsymbol{h} = \frac{1}{2^j}\mathbb{1}(\mathcal{W})$. Therefore, $\boldsymbol{h} \in 2\hat{\mathcal{U}}_{j+1}$. Consequently, we get $\boldsymbol{h} \in 2\mathcal{U}$ from Claim 5.2.4. Hence, by feasibility of the optimal solution, we get

$$2\boldsymbol{A}\boldsymbol{x}^* + 2\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{h}/2) \ge \boldsymbol{h}.$$

Furthermore, for all $i \in \mathcal{W}$, we have $i \notin \mathcal{J}_{1j}$ and therefore $4(\boldsymbol{A}\boldsymbol{x}^*)_i < \frac{1}{2^j}$ which implies

$$4(\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{h}/2))_i \ge 2h_i - \frac{1}{2^j} = h_i.$$

i.e.,

$$4\boldsymbol{B}\boldsymbol{y}^*(\boldsymbol{h}/2) \ge \boldsymbol{h}$$

This means that $4\boldsymbol{y}^*(\boldsymbol{h}/2)$ is a feasible solution for the covering problem (5.10) with requirement

$\boldsymbol{h}$, therefore,

$$z\left(\frac{1}{2^j}\mathbb{1}(\mathcal{W})\right) = z(\boldsymbol{h}) \leq 4\boldsymbol{d}^T\boldsymbol{y}^*(\boldsymbol{h}/2) \leq 4\text{OPT}_2 \leq 4\text{OPT} = \frac{1}{2^j}\gamma.$$

Therefore,

$$z\left(\mathbb{1}(\mathcal{W})\right) \leq \gamma.$$

i.e., condition 2 of Lemma 5.3.9 is also satisfied. Therefore from Lemma 5.3.9, we have

$$z\left(\frac{1}{2^j}\mathbb{1}(\mathcal{J}_{2j})\right) \leq \frac{\gamma}{2^j}\frac{\alpha}{2} = 2\alpha \cdot \text{OPT}.$$

Denote $\boldsymbol{y}_j$ an optimal solution corresponding to $z\left(\frac{1}{2^j}\mathbb{1}(\mathcal{J}_{2j})\right)$, i.e., $\boldsymbol{d}^T\boldsymbol{y}_j = z\left(\frac{1}{2^j}\mathbb{1}(\mathcal{J}_{2j})\right)$. Hence, we have $(4\boldsymbol{x}^*, \boldsymbol{y}_j)$ is a feasible solution for the static problem to cover $\sum_{i\in\mathcal{I}_j}\frac{1}{2^j}\boldsymbol{e}_i$. Therefore,

$$\sum_{j=0}^{j_{\max}}\left(4\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}_j\right) \geq \sum_{j=0}^{j_{\max}}\sum_{i\in\mathcal{I}_j}\frac{1}{2^j}\boldsymbol{e}_i = \sum_{i=1}^{m}\theta_i\boldsymbol{e}_i$$

Moreover,

$$\sum_{j=0}^{j_{\max}}(4\boldsymbol{c}^T\boldsymbol{x}^* + \boldsymbol{d}^T\boldsymbol{y}_j) \leq (j_{\max}+1)(4\text{OPT}_1 + 2\alpha\cdot\text{OPT}) = O(\log m\cdot\alpha)\cdot\text{OPT},$$

Hence, the cost of the static problem (5.7) is bounded by $O(\alpha\log m)\cdot\text{OPT}$

□

The proof of Theorem 5.3.1 follows directly from combining Lemma 5.3.8 and Lemma 5.3.7.

Chapter 6: Piecewise static policies

6.1 Introduction

In this chapter, we consider two-stage adjustable robust linear optimization problems under packing uncertain constraints and study the performance of piecewise static policies. These are a generalization of static policies where we divide the uncertainty set into several pieces and specify a static solution for each piece. We show that in general there is no piecewise static policy with a polynomial number of pieces that has a significantly better performance than an optimal static policy. This is quite surprising as piecewise static policies are significantly more general than static policies. More specifically, recall two-stage adjustable robust problem with packing constraints (1.7)

$$
\begin{aligned}
z_{\mathsf{AR}}(\mathcal{U}) = \max \;\; & \boldsymbol{c}^T \boldsymbol{x} + \min_{\boldsymbol{B} \in \mathcal{U}} \max_{\boldsymbol{y}(\boldsymbol{B})} \boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{B}) \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{B}) \leq \boldsymbol{h} \\
& \boldsymbol{x} \in \mathbb{R}^n_+, \; \boldsymbol{y}(\boldsymbol{B}) \in \mathbb{R}^n_+.
\end{aligned}
$$

As mentioned before, following Bertsimas et al. [43], we can assume without loss of generality that $\mathcal{U}$ is *down-monotone* (A set $\mathcal{S} \subseteq \mathbb{R}^n_+$ is *down-monotone* if $\boldsymbol{s} \in \mathcal{S}$, $\boldsymbol{t} \in \mathbb{R}^n_+$ and $\boldsymbol{t} \leq \boldsymbol{s}$ implies $\boldsymbol{t} \in \mathcal{S}$). The above formulation models many interesting applications including revenue management and resource allocation problems with uncertain demand. For instance, in a resource allocation application, the right hand side $\boldsymbol{h}$ can model the fixed resource capacities and the uncertain coefficients in $\boldsymbol{B}$ model the uncertain requirements of resources for demand. The goal is to find an optimal allocation of resources that maximizes the worst case profit (see Wiesemann [44]). In general, it is intractable to compute an adjustable robust solution for (1.7). In fact, Awasthi et al. [48] show that the two-stage adjustable robust problem (1.7) is $\Omega(\log n)$-hard to approximate if the uncertainty set of constraint coefficients belongs to the non-negative orthant. In other words, there is no polynomial time algorithm that approximates the optimal adjustable solution within a factor better than $\log n$. Therefore, the goal is to construct approximate policies with good performance. A static solution approach, where we give a single solution feasible for all scenarios, has been widely studied in the literature. Recall the static robust optimization problem (1.8) to approximate (1.7) is given by

$$
\begin{aligned}
z_{\mathsf{Rob}}(\mathcal{U}) = \max \ & \boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{d}^T \boldsymbol{y} \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \leq \boldsymbol{h} \ \ \forall \boldsymbol{B} \in \mathcal{U} \\
& \boldsymbol{x} \in \mathbb{R}^n_+, \ \boldsymbol{y} \in \mathbb{R}^n_+.
\end{aligned}
$$

As we mention earlier, an optimal static solution can be computed efficiently for large class of problems (see Bertsimas et al. [15], Ben-Tal et al. [16]). Ben-Tal and Nemirovski [9] show that a static solution is optimal for (1.7) if the uncertainty set is constraint-wise where each constraint is selected independently from a compact convex set $\mathcal{U}_i$ (i.e. $\mathcal{U}$ is a Cartesian product of $\mathcal{U}_i$, $i = 1, \ldots, m$). Bertsimas et al.[43] generalize the result of [9] and show that a static solution is near optimal for several interesting families of $\mathcal{U}$. In particular, they give a tight characterization on the performance of the static solution related to the measure of non-convexity of a transformation of the uncertainty set $\mathcal{U}$. While a static solution provides a good approximation in many cases, it can be as bad as a factor m away from the optimal adjustable solution in general.

In this chapter, we consider the piecewise static solution approach for (1.7). A piecewise static policy (also referred to as *finite K-adaptability*) is a generalization of the static policy where the uncertainty set is divided into several pieces and we specify a static policy for each piece [39][31]. In particular, we consider a piecewise policy with p pieces (or subsets): $\mathcal{U}_1, \ldots, \mathcal{U}_p$ of $\mathcal{U}$ such that

$$
\mathcal{U} = \bigcup_{1 \leq i \leq p} \mathcal{U}_i,
$$

where each $\mathcal{U}_i$ is convex, compact and down-monotone uncertainty subset. Note that $\mathcal{U}_i$ are not necessarily disjoint. We can formulate the two-stage piecewise robust linear optimization problem as in 1.9, i.e.,

$$
\begin{aligned}
z_{\mathsf{PR}}(\mathcal{U}_1,\ldots,\mathcal{U}_p) = \max \;\; & \boldsymbol{c}^T\boldsymbol{x} + \min(\boldsymbol{d}^T\boldsymbol{y}_1, \boldsymbol{d}^T\boldsymbol{y}_2, \ldots, \boldsymbol{d}^T\boldsymbol{y}_p) \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}_i\boldsymbol{y}_i \leq \boldsymbol{h} \qquad \forall i \in [p]\ ,\ \forall \boldsymbol{B}_i \in \mathcal{U}_i \\
& \boldsymbol{x} \in \mathbb{R}^n_+,\ \boldsymbol{y}_i \in \mathbb{R}^n_+ \;\; \forall i \in [p].
\end{aligned}
$$

We show that the performance of the optimal piecewise static policy for given pieces is related to the maximum of the measures of non-convexity of transformations of the pieces $\mathcal{U}_i$; thereby extending the bound in [43] for piecewise static policies. Note that if the pieces $\mathcal{U}_i$ are given explicitly, we can efficiently compute an optimal piecewise static policy provide we can solve linear optimization over each $\mathcal{U}_i$ efficiently. However, one of the main challenges in designing a good piecewise static policy, is to construct good pieces of the uncertainty set. In fact, Bertsimas and Caramanis [39] show that it is NP-hard to construct the optimal pieces for piecewise policies with only two pieces for two-stage robust linear programs in general.

Our main contribution in this chapter is to show that even if we ignore the computational complexity of computing optimal pieces, surprisingly the performance of piecewise static policies with a polynomial number of pieces is not significantly better than a static policy in general. In particular, we show that there is no piecewise static policy with polynomial number of pieces that gives an approximation bound better than $O(m^{1-\epsilon})$ for any $\epsilon > 0$ for general uncertainty sets $\mathcal{U} \subseteq R^{m\times n}_+$ where the approximation bound for the static policy is m. We prove this by constructing a family of instances of $\mathcal{U}$ for any $\epsilon > 0$, such that the performance of the static policy is m and the performance of any piecewise policy with polynomial number of pieces is $\Omega(m^{1-\epsilon})$. Our proof is based on a combinatorial argument and structural results about piecewise static policies.

The rest of the chapter is organized as follows. We present the preliminaries in Section 6.2. In Section 6.3, we present the structural results for piecewise static policies. Finally, we present the lower bound on the performance of piecewise static policies in Section 6.4.

6.2 Preliminaries: Static policies

In this section, we present some preliminaries and definitions for our results. As we mention earlier, Bertsimas et al. [43] give a tight characterization on the performance of a static solution as compared to the optimal adjustable solution for problem (1.7). They relate this performance to the measure of non-convexity of a transformation of the uncertainty set. We first introduce the following definitions.

Definition 6.2.1. (Transformation $T(U,\cdot)$). For any $\boldsymbol{h} > \boldsymbol{0}$ and convex compact full-dimensional down-monotone set $\mathcal{U} \subseteq R_+^{m\times n}$, we define the following transformation:

$$T(\mathcal{U},\boldsymbol{h}) = \{\boldsymbol{B}^T\boldsymbol{\mu} \mid \boldsymbol{h}^T\boldsymbol{\mu} = 1, \boldsymbol{B} \in \mathcal{U}, \boldsymbol{\mu} \geq \boldsymbol{0}\}.$$

Definition 6.2.2. (Measure of non-convexity). For any down-monotone compact set $\mathcal{S} \subseteq R_+^n$, the measure of non-convexity $\kappa(\mathcal{S})$ is defined as follows:

$$\kappa(\mathcal{S}) = \min\{\alpha \mid \mathsf{conv}(\mathcal{S}) \subseteq \alpha \cdot \mathcal{S}\}.$$

Definition 6.2.3. For any convex compact full-dimensional down-monotone set $\mathcal{U}$, let,

$$\rho(\mathcal{U}) = \max_{\boldsymbol{h}>\boldsymbol{0}} \kappa(T(\mathcal{U},\boldsymbol{h})).$$

For any $\mathcal{U} \subset \mathbb{R}_+^{m\times n}$, Bertsimas et al.[43] give the following characterization of $\mathsf{conv}(T(U,\cdot))$.

Lemma 6.2.4 (Bertsimas et al. [43]). *For any* $\boldsymbol{h} > \boldsymbol{0}$,

$$\mathsf{conv}(T(\mathcal{U},\boldsymbol{h})) = \mathsf{conv}\left(\bigcup_{1\leq i\leq m}\left\{\frac{1}{h_i}\boldsymbol{B}^T\boldsymbol{e}_i \;\middle|\; \boldsymbol{B} \in \mathcal{U}\right\}\right).$$

Consider the following one-stage adjustable robust problem, $\Pi^I_{\mathsf{AR}}(\mathcal{U},\boldsymbol{h})$, corresponding to (1.7).

$$z^I_{\mathsf{AR}}(\mathcal{U},\boldsymbol{h}) = \min_{\boldsymbol{B}\in\mathcal{U}} \max_{\boldsymbol{y}\geq\boldsymbol{0}} \{\boldsymbol{d}^T\boldsymbol{y} \mid \boldsymbol{B}\boldsymbol{y} \leq \boldsymbol{h}\}. \tag{6.1}$$

The one-stage problem is related to the separation problem for the two-stage adjustable robust optimization problem (1.7). Similarly, we can consider the following one-stage robust problem, $\Pi^I_{\mathsf{Rob}}(\mathcal{U},\boldsymbol{h})$, corresponding to (1.8).

$$z^I_{\mathsf{Rob}}(\mathcal{U},\boldsymbol{h}) = \max_{\boldsymbol{y}\geq\boldsymbol{0}} \{\boldsymbol{d}^T\boldsymbol{y} \mid \boldsymbol{B}\boldsymbol{y} \leq \boldsymbol{h}\ \forall \boldsymbol{B} \in \mathcal{U}\}. \tag{6.2}$$

Bertsimas et al. [43] give the following reformulations of (6.1) and (6.2).

Lemma 6.2.5 (Bertsimas et al. [43])**.** $\Pi^I_{\mathsf{AR}}(\mathcal{U},\boldsymbol{h})$ (6.1) *can be reformulated as*

$$z^I_{\mathsf{AR}}(\mathcal{U},\boldsymbol{h}) = \min\ \{\lambda \mid \lambda\boldsymbol{b} \geq \boldsymbol{d}, \boldsymbol{b} \in T(\mathcal{U},\boldsymbol{h})\}.$$

Lemma 6.2.6 (Bertsimas et al. [43])**.** $\Pi^I_{\mathsf{Rob}}(\mathcal{U},\boldsymbol{h})$ *can be formulated as*

$$z^I_{\mathsf{Rob}}(\mathcal{U},\boldsymbol{h}) = \min\{\lambda \mid \lambda\boldsymbol{b} \geq \boldsymbol{d},\ \boldsymbol{b} \in \mathsf{conv}(T(\mathcal{U},\boldsymbol{h})\}.$$

Furthermore, they show that

$$z_{\mathsf{Rob}}(\mathcal{U}) \leq z_{\mathsf{AR}}(\mathcal{U}) \leq \rho(\mathcal{U}) \cdot z_{\mathsf{Rob}}(\mathcal{U}),$$

where $\rho(\mathcal{U})$ is the tight bound that characterizes the performance of the static policy. Note that $\rho(\mathcal{U})$ can be as bad as m in general. The worst case instance for $\rho(\mathcal{U})$ is the diagonal uncertainty set

$$\mathcal{U} = \left\{\mathsf{diag}(\boldsymbol{x}) \;\middle|\; \sum_{i=1}^{m} x_i \leq 1,\ \boldsymbol{x} \geq \boldsymbol{0}\right\}. \tag{6.3}$$

For this example of uncertainty set we have, $z_{\mathsf{AR}}(\mathcal{U}) = m \cdot z_{\mathsf{Rob}}(\mathcal{U})$. We refer the reader to Bertsimas et al. [43] for more details.

6.3 Structural results on piecewise static policy

In this section, we introduce the piecewise static policies for the two-stage adjustable robust optimization problem (1.7) and study the structural properties and performance of these policies. We first introduce the following definition.

Definition 6.3.1. (Convex cover) Let $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_p$ subsets of $\mathcal{U}$ such that $\mathcal{U}_i$ is convex, compact and down-monotone set. We say that $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_p$ is a convex cover of $\mathcal{U}$ if $\mathcal{U} = \mathcal{U}_1 \cup \mathcal{U}_2 \ldots \cup \mathcal{U}_p$.

Note that different pieces are not necessarily disjoint. We only require that the union of pieces covers $\mathcal{U}$.

6.3.1 Performance of piecewise static policy

Let $\mathcal{U} = \mathcal{U}_1 \cup \mathcal{U}_2 \ldots \cup \mathcal{U}_p$ be a convex cover of $\mathcal{U}$. We relate the performance of the optimal piecewise static solution to the maximum of the measures of non-convexity of the transformations $T(\mathcal{U}_i, \cdot)$. Consider the following reformulation of the two-stage piecewise static robust linear optimization problem (1.9).

$$
\begin{aligned}
z_{\mathsf{PR}}(\mathcal{U}_1, \ldots, \mathcal{U}_p) = \max \ & \boldsymbol{c}^T \boldsymbol{x} + z \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}_i \boldsymbol{y}_i \leq \boldsymbol{h} \qquad \forall i \in [p]\ ,\ \forall \boldsymbol{B}_i \in \mathcal{U}_i \\
& z \leq \boldsymbol{d}^T \boldsymbol{y}_i \qquad \forall i \in [p] \\
& \boldsymbol{x} \in \mathbb{R}^n_+,\ \boldsymbol{y}_i \in \mathbb{R}^n_+\ \forall i \in [p],\ z \in \mathbb{R}.
\end{aligned}
\tag{6.4}
$$

We can compute the solution of this problem efficiently if the number of pieces is small and linear optimization is efficient over each piece.

Let $(\boldsymbol{x}^*, (\boldsymbol{y}_1^*, \boldsymbol{y}_2^*, \ldots, \boldsymbol{y}_p^*))$ be an optimal solution of (6.4). Then $(\boldsymbol{x}^*, \boldsymbol{y}(\boldsymbol{B}))$, where $\boldsymbol{y}(\boldsymbol{B}) = \boldsymbol{y}_i^*$ if $\boldsymbol{B} \in \mathcal{U}_i$, is a feasible solution for the adjustable problem (1.7). Therefore,

$$
z_{\mathsf{PR}}(\mathcal{U}_1, \ldots, \mathcal{U}_p) \leq z_{\mathsf{AR}}(\mathcal{U}). \tag{6.5}
$$

To compute an upper bound for $z_{\mathsf{AR}}(\mathcal{U})$ in terms of $z_{\mathsf{PR}}(\mathcal{U}_1,\ldots,\mathcal{U}_p)$, consider the following one stage piecewise static problem $\Pi^I_{\mathsf{PR}}((\mathcal{U}_1,\ldots,\mathcal{U}_p),\boldsymbol{h})$:

$$z^I_{\mathsf{PR}}((\mathcal{U}_1,..,\mathcal{U}_p),\boldsymbol{h}) = \max_{\boldsymbol{y}_i\geq\boldsymbol{0}} \{z \mid \boldsymbol{B}_i\boldsymbol{y}_i \leq \boldsymbol{h}\ \forall \boldsymbol{B}_i\in\mathcal{U}_i,\ z\leq \boldsymbol{d}^T\boldsymbol{y}_i,\ \forall i\in[p]\} \tag{6.6}$$

Lemma 6.3.2. *For the one stage piecewise static problem* $\Pi^I_{\mathsf{PR}}((\mathcal{U}_1,\ldots,\mathcal{U}_p),\boldsymbol{h})$,

$$z^I_{\mathsf{PR}}((\mathcal{U}_1,\ldots,\mathcal{U}_p),\boldsymbol{h}) = \min_{1\leq i\leq p} z^I_{\mathsf{Rob}}(U_i,\boldsymbol{h}).$$

Lemma 6.3.2 follows directly from (6.6). The following theorem relates the performance of a piecewise static solution to the measures of non-convexity of $T(\mathcal{U}_i,\boldsymbol{h})$.

Theorem 6.3.3. *For any convex cover of* $\mathcal{U}$ *such that* $\mathcal{U}=\mathcal{U}_1\cup\mathcal{U}_2\ldots\cup\mathcal{U}_p$, *we have,*

$$z_{\mathsf{AR}}(\mathcal{U}) \leq \max(\rho(\mathcal{U}_1),\ldots,\rho(\mathcal{U}_p))\cdot z_{\mathsf{PR}}(\mathcal{U}_1,\ldots,\mathcal{U}_p).$$

Furthermore, the bound is tight.

Proof. Denote $\hat{\lambda}_\ell$, $\hat{\boldsymbol{b}}_\ell \in \mathsf{conv}(T(\mathcal{U}_\ell,\boldsymbol{h}))$ the solutions of the one stage piecewise static problem $\Pi^I_{\mathsf{PR}}((\mathcal{U}_1,\ldots,\mathcal{U}_p),\boldsymbol{h})$ under the formulations of Lemma 6.3.2 and Lemma 6.2.6, where $\ell\in[p]$. We have $z^I_{\mathsf{PR}}((\mathcal{U}_1,\ldots,\mathcal{U}_p),\boldsymbol{h}) = \hat{\lambda}_\ell$ and $\hat{\lambda}_\ell\hat{\boldsymbol{b}}_\ell \geq \boldsymbol{d}$, i.e.

$$\kappa_\ell\hat{\lambda}_\ell\cdot\frac{\hat{\boldsymbol{b}}_\ell}{\kappa_\ell} \geq \boldsymbol{d}$$

where $\kappa_\ell = \kappa(T(\mathcal{U}_\ell,\boldsymbol{h}))$. Since,

$$\frac{\hat{\boldsymbol{b}}_\ell}{\kappa_\ell} \in T(\mathcal{U}_\ell,\boldsymbol{h}) \subseteq T(\mathcal{U},\boldsymbol{h}),$$

then $(\kappa_\ell\hat{\lambda}_\ell)$ is a feasible solution for $\Pi^I_{\mathsf{AR}}(\mathcal{U},\boldsymbol{h})$ under the formulation of Lemma 6.2.5, i.e. $\kappa_\ell\hat{\lambda}_\ell \geq z^I_{\mathsf{AR}}(\mathcal{U},\boldsymbol{h})$.

Moreover, we know that $\max(\rho(\mathcal{U}_1),\ldots,\rho(\mathcal{U}_p)) \geq \kappa_\ell$. Then,

$$\max(\rho(\mathcal{U}_1),\ldots,\rho(\mathcal{U}_p)) \cdot z^I_{\mathsf{PR}}((\mathcal{U}_1,\ldots,\mathcal{U}_p),\boldsymbol{h}) \geq z^I_{\mathsf{AR}}(\mathcal{U},\boldsymbol{h}).$$

Therefore,

$$\begin{aligned} z_{\mathsf{AR}}(\mathcal{U},\boldsymbol{h}) &= \boldsymbol{c}^T\boldsymbol{x}^* + z^I_{\mathsf{AR}}(\mathcal{U},\boldsymbol{h}-\boldsymbol{A}\boldsymbol{x}^*) \\ &\leq \boldsymbol{c}^T\boldsymbol{x}^* + \max(\rho(\mathcal{U}_1),\ldots,\rho(\mathcal{U}_p)) \cdot z^I_{\mathsf{PR}}((\mathcal{U}_1,\ldots,\mathcal{U}_p),\boldsymbol{h}-\boldsymbol{A}\boldsymbol{x}^*) \\ &\leq \max(\rho(\mathcal{U}_1),\ldots,\rho(\mathcal{U}_p)) \cdot (\boldsymbol{c}^T\boldsymbol{x}^* + z^I_{\mathsf{PR}}((\mathcal{U}_1,\ldots,\mathcal{U}_p),\boldsymbol{h}-\boldsymbol{A}\boldsymbol{x}^*)) \\ &\leq \max(\rho(\mathcal{U}_1),\ldots,\rho(\mathcal{U}_p)) \cdot z_{\mathsf{PR}}(\mathcal{U}_1,\ldots,\mathcal{U}_p). \end{aligned}$$

The last inequality follows from the definition (6.6) of the one stage piecewise static problem. The tightness of the bound follows from the tightness of the bound for static policies [43]. □

6.3.2 Examples of piecewise static policies

We present several examples to illustrate the performance bound for piecewise static policies. In particular, we consider the diagonal uncertainty set defined in (6.3) for which the performance of static policies is the worst possible as compared to the optimal fully adjustable solution. We first show that without loss of generality, we can consider pieces of the following form for any convex cover of $\mathcal{U}$ (6.3).

$$\mathcal{V}(\tau_1,\tau_2,\ldots,\tau_m) = \left\{\mathsf{diag}(\boldsymbol{x}) \;\middle|\; \sum_{j=1}^m x_j \leq 1,\ 0 \leq x_j \leq \tau_j\ \forall j \in [m]\right\}. \tag{6.7}$$

In particular, we have the following structural lemma.

Lemma 6.3.4 (Structure of Piecewise static policies)**.** *Let $\mathcal{U} = \mathcal{U}_1 \cup \mathcal{U}_2 \ldots \cup \mathcal{U}_p$ a convex cover of the diagonal uncertainty set* (6.3)*. For all $i \in [p]$ we define, $\mathcal{V}_i = \mathcal{V}(\tau_{i1},\tau_{i2},\ldots,\tau_{im})$, where for* *all $i \in [p]$ and $j \in [m]$,*

$$\tau_{ij} = \max_{\mathsf{diag}(\boldsymbol{x}) \in \mathcal{U}_i} \boldsymbol{e}_j^T \boldsymbol{x}.$$

Then, $\forall i \in [p]$, $\mathcal{U}_i \subseteq \mathcal{V}_i \subseteq \mathcal{U}$ and $\kappa(T(\mathcal{V}_i, \boldsymbol{h})) \leq \kappa(T(\mathcal{U}_i, \boldsymbol{h}))$.

Proof. Let $i \in [p]$. We have $\forall$ $\mathsf{diag}(\boldsymbol{x}) \in \mathcal{U}_i$, $x_j \leq \tau_{ij}$ for $j = 1, \ldots, m$. Then, $\mathcal{U}_i \subseteq \mathcal{V}_i \subseteq \mathcal{U}$. Now, we will show that for all $i \in [p]$, $\mathsf{conv}(T(\mathcal{U}_i, \boldsymbol{h})) = \mathsf{conv}(T(\mathcal{V}_i, \boldsymbol{h}))$. First, since $\mathcal{U}_i \subseteq \mathcal{V}_i$, clearly, $\mathsf{conv}(T(\mathcal{U}_i, \boldsymbol{h})) \subseteq \mathsf{conv}(T(\mathcal{V}_i, \boldsymbol{h}))$. Consider any $\boldsymbol{b} \in T(\mathcal{V}_i, \boldsymbol{h})$. Then,

$$\boldsymbol{b} = \mathsf{diag}(\boldsymbol{x})^T \boldsymbol{\mu},$$

where $\sum_{k=1}^m \mu_k h_k = 1$ and $\mathsf{diag}(\boldsymbol{x}) \in \mathcal{V}_i$. Therefore,

$$\boldsymbol{b} = \sum_{k=1}^m \mu_k h_k \cdot \frac{x_k}{h_k} \boldsymbol{e}_k$$

For all $k \in [m]$, we have $x_k \leq \tau_{ik}$ and we know that $\mathcal{U}_i$ is *down-monotone*. Therefore, $x_k \boldsymbol{e}_k \in \mathcal{U}_i$ and $\frac{x_k}{h_k} \boldsymbol{e}_k \in T(\mathcal{U}_i, \boldsymbol{h})$. Hence $\boldsymbol{b} \in \mathsf{conv}(T(\mathcal{U}_i, \boldsymbol{h}))$ and $\mathsf{conv}(T(\mathcal{V}_i, \boldsymbol{h})) \subseteq \mathsf{conv}(T(\mathcal{U}_i, \boldsymbol{h}))$. Therefore,

$$\begin{aligned} \mathsf{conv}(T(\mathcal{V}_i, \boldsymbol{h})) = \mathsf{conv}(T(\mathcal{U}_i, \boldsymbol{h})) &\subseteq \kappa(T(\mathcal{U}_i, \boldsymbol{h})) \cdot T(U_i, h) \\ &\subseteq \kappa(T(\mathcal{U}_i, \boldsymbol{h})) \cdot T(\mathcal{V}_i, \boldsymbol{h}), \end{aligned}$$

which implies $\kappa(T(\mathcal{V}_i, \boldsymbol{h})) \leq \kappa(T(\mathcal{U}_i, \boldsymbol{h}))$. □

In the following lemma, we show that we can compute the measure of non-convexity of $T(\mathcal{V}(\tau_1, \tau_2, \ldots, \tau_m), \boldsymbol{h})$ where $\mathcal{V}(\tau_1, \tau_2, \ldots, \tau_m)$ is defined in (6.7).

Lemma 6.3.5. *Let,*

$$\mathcal{U} = \mathcal{V}(\tau_1, \tau_2, \ldots, \tau_m)$$

where $\mathcal{V}(\tau_1, \tau_2, \ldots, \tau_m)$ is defined in (6.7) *such that $\forall i \in [m]$, $0 \leq \tau_i \leq 1$ and $\sum_{i=1}^m \tau_i \geq 1$. Then* *for all* $\boldsymbol{h} > \mathbf{0}$,

$$\kappa(T(\mathcal{U},\boldsymbol{h})) = \sum_{i=1}^{m} \tau_i.$$

The proof of Lemma 6.3.5 is presented in Appendix D.1. We now present two examples of convex covers of the diagonal uncertainty set $\mathcal{U}$ (6.3) and give the performance of the corresponding piecewise static policy for each example.

Example 1. For all $j = 1, \ldots, m$ let,

$$\mathcal{U}_j = \left\{ \mathsf{diag}(\boldsymbol{x}) \;\middle|\; \sum_{i=1}^{m} x_i \leq 1,\ 0 \leq x_j \leq \frac{1}{m} \right\}.$$

Note that $\underset{1\leq j\leq m}{\cup} \mathcal{U}_j$ is a convex cover of $\mathcal{U}$ with m number of pieces. From Lemma 6.3.5, we have the following.

Propostion 6.3.6. For the cover defined in Example 1, the performance of piecewise static policy is

$$\rho = m - 1 + \frac{1}{m}.$$

Example 2. Let $\mathcal{S}_m$ be the set of permutations in $\{1, 2, \ldots, m\}$ and let $\boldsymbol{\tau} = (1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{m})$. For all $\sigma \in \mathcal{S}_m$ let,

$$\mathcal{U}_\sigma = \left\{ \mathsf{diag}(\boldsymbol{x}) \;\middle|\; 0 \leq x_i \leq \tau_{\sigma(i)}\ \forall i \in [m]\ ,\ \sum_{i=1}^{m} x_i \leq 1 \right\},$$

Note that $\underset{\sigma\in\mathcal{S}_m}{\cup} U_\sigma$ is a convex cover of $\mathcal{U}$ with $m!$ number of pieces. From Lemma 6.3.5, we have the following.

Propostion 6.3.7. For the cover defined in Example 2, the performance of piecewise static policy is

$$\rho = \sum_{i=1}^{m} \frac{1}{i} = O(\log(m)).$$

We would like to note that for the cover in Example 1, the number of pieces is polynomial and the performance bound for the piecewise static policy is $\Omega(m)$ which is the same order as the approximation bound for static policies. For Example 2, the performance bound for the piecewise static policy is $O(\log m)$ which is significantly better. However the number of pieces is exponential. Since it is difficult to compute a piecewise static policy with exponentially many pieces, it motivates us to consider the problem of finding piecewise static policies with a polynomial number of pieces that have a significantly better performance than the static policy.

6.4 Lower bound for polynomial pieces

In this section, we show that, surprisingly there is no piecewise static policy with polynomial number of pieces that gives an approximation bound significantly better than the static policies in general. In particular, we consider the diagonal uncertainty set (6.3). Bertsimas et al. [43] present family of instances where $z_{\mathsf{AR}}(\mathcal{U}) = m \cdot z_{\mathsf{Rob}}(\mathcal{U})$ for the uncertainty set (6.3). We show that for any fixed $\epsilon > 0$, there is no piecewise static policy with polynomial number of pieces with approximation bound as $O(m^{1-\epsilon})$. Our proof is based on a combinatorial argument that exploits the structural result for piecewise policies for (6.3) derived in the previous section. We have the following theorem.

Theorem 6.4.1 (Main result)**.** *For any given $0 < \epsilon < 1$ and $k \in \mathbb{N}$, there are instances of uncertainty set $\mathcal{U} \subset \mathbb{R}_+^{m\times n}$ with sufficiently large m such that for any convex cover $(\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_p)$ of $\mathcal{U}$ with $p \leq (\max(m,n))^k$ pieces,*

$$\max(\rho(\mathcal{U}_1), \ldots, \rho(\mathcal{U}_p)) > m^{1-\epsilon}.$$

Proof. Consider the diagonal uncertainty set $\mathcal{U} \subset \mathbb{R}_+^{m\times m}$ defined in (6.3) for m sufficiently large. Consider $(\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_p)$ a convex cover of $\mathcal{U}$ such that $p \leq m^k$. We can assume without loss of generality $p = m^k$. Suppose for the sake of contradiction,

$$\max(\rho(\mathcal{U}_1), \ldots, \rho(\mathcal{U}_p)) \leq m^{1-\epsilon}. \tag{6.8}$$

From Lemma 6.3.4, it is sufficient to consider $\mathcal{U}_i$ of the following form for all $i \in [p]$:

$$\mathcal{U}_i = \left\{ \mathsf{diag}(\boldsymbol{x}) \;\middle|\; \sum_{i=1}^{m} x_i \leq 1,\ 0 \leq x_j \leq \tau_{ij} \ \ \forall j \in [m] \right\}$$

From Lemma 6.3.5, for all $i \in [p]$, $\forall \boldsymbol{h} > \mathbf{0}$,

$$\kappa(T(\mathcal{U}_i, \boldsymbol{h})) = \sum_{j=1}^{m} \tau_{ij} \leq m^{1-\epsilon}$$

where the last inequality follows from the assumption (6.8). Let

$$\beta = \left\lfloor \frac{1}{\epsilon} \right\rfloor .$$

We define the following discrete set

$$\mathcal{W} = \left\{ \mathsf{diag}\left(\frac{a_1}{\gamma}, \ldots, \frac{a_m}{\gamma} \right) \;\middle|\; \sum_{i=1}^{m} a_i = \gamma,\ a_i \in \{0,1\}, \forall i \in [m] \right\},$$

where $\gamma = \beta k + k + 1$. Note that $\mathcal{W}$ is a discrete subset of $\mathcal{U}$ with cardinality

$$|\mathcal{W}| = \binom{m}{\gamma} = \binom{m}{\beta k + k + 1} = \Theta\left(m^{\beta k + k + 1}\right).$$

We have

$$\mathcal{W} \subseteq \mathcal{U} = \underset{1 \leq i \leq p}{\cup} \mathcal{U}_i .$$

Hence there exists $1 \leq \ell \leq m^k$ such that $\mathcal{U}_\ell$ contains at least $\frac{|\mathcal{W}|}{m^k}$ elements of $\mathcal{W}$. In particular, there exists $\hat{\mathcal{W}} \subseteq \mathcal{W}$ such that $\hat{\mathcal{W}} \subseteq \mathcal{U}_\ell$ and

$$|\hat{\mathcal{W}}| \geq \frac{|\mathcal{W}|}{m^k} = \Theta(m^{\beta k + 1}). \tag{6.9}$$

Then $\forall j \in [m]$ and $\forall \boldsymbol{a} \in \hat{\mathcal{W}}$,

$$\frac{a_j}{\beta k + k + 1} \leq \tau_{\ell j}.$$

Therefore, $\forall j \in [m]$,

$$\frac{\max_{\boldsymbol{a} \in \hat{\mathcal{W}}} \boldsymbol{e}_j^T \boldsymbol{a}}{\beta k + k + 1} \leq \tau_{\ell j},$$

which implies

$$\begin{aligned}
\sum_{j=1}^{m} \max_{\boldsymbol{a} \in \hat{\mathcal{W}}} \boldsymbol{e}_j^T \boldsymbol{a} &\leq (\beta k + k + 1) \sum_{j=1}^{m} \tau_{\ell j} \\
&\leq (\beta k + k + 1) m^{1-\epsilon} \\
&< (\beta k + k + 1) m^{\frac{\beta}{\beta+1}},
\end{aligned}$$

where the last inequality follows from $\frac{1}{\beta+1} < \epsilon$. Denote $t = (\beta k + k + 1) m^{\frac{\beta}{\beta+1}}$ and

$$\mathcal{S} = \{j \in [m] \mid \exists \boldsymbol{a} \in \hat{\mathcal{W}},\ a_j = 1\}.$$

Then, $|\mathcal{S}| \leq \lfloor t \rfloor$. We have,

$$\hat{\mathcal{W}} \subseteq \left\{ \mathsf{diag}\left(\frac{a_1}{\gamma}, \ldots, \frac{a_m}{\gamma} \right) \;\middle|\; \sum_{i \in \mathcal{S}} a_i = \gamma,\ a_i \in \{0, 1\}, \forall i \in [m] \right\}.$$

Therefore,

$$|\hat{\mathcal{W}}| \leq \binom{|\mathcal{S}|}{\gamma} = \Theta\left(|\mathcal{S}|^{\gamma}\right).$$

We have,

$$
\begin{aligned}
|\mathcal{S}|^{\gamma} = |\mathcal{S}|^{\beta k+k+1} &\leq \lfloor t \rfloor^{\beta k+k+1} \\
&\leq t^{\beta k+k+1} \\
&= \left((\beta k + k + 1) m^{\frac{\beta}{\beta+1}} \right)^{(\beta k+k+1)} \\
&= (\beta k + k + 1)^{(\beta k+k+1)} \cdot m^{\beta k + \frac{\beta}{\beta+1}}
\end{aligned}
$$

Then,

$$
|\hat{\mathcal{W}}| \leq \Theta\left(m^{\beta k + \frac{\beta}{\beta+1}} \right).
$$

On the other hand, $|\hat{\mathcal{W}}| \geq \Theta\left(m^{\beta k+1} \right)$ (6.9) which is a contradiction for m sufficiently large. □

The above theorem implies that if we restrict to piecewise policies with a polynomial number of pieces, we can not get significantly better policies than static in general. This is quite surprising since piecewise static policies are more general than a single static solution.

Conclusion

This thesis focuses on some fundamental questions in the theory and foundations of robust optimization. At a high level, the thesis addresses two challenges. The first one is to bridge the gap between the empirical and theoretical performance of simple policies in robust optimization such as affine policies. The second one is to design new polices that are tractable, scalable and significantly improve over affine and static policies.

In fact, while the worst-case performance of affine policies can be arbitrarily bad, the empirical performance is observed to be near-optimal on both synthetic and real data. We present a fine-grained analysis of affine policies that addresses this stark contrast between theory and practice in two different ways. First, we introduce in Chapter 2 a probabilistic approach to analyze the performance of affine policies on randomly generated instances of two-stage robust optimization. We show that with high probability affine policies give a good approximation for a wide range of instances drawn from a large class of distributions; thereby, providing a theoretical justification of the observed empirical performance. It is an interesting question to extend this probabilistic analysis to other classes of dynamic robust optimization problems.

Second, we study the performance of affine policies for an important class of uncertainty sets widely used in practice, namely budget of uncertainty. In particular, we show in Chapter 3 that affine policies give the optimal approximation for two-stage adjustable problem with covering constraints under budget of uncertainty sets which confirms the power of these policies and explains the good empirical performance under this widely used class of uncertainty sets. We also provide strong theoretical bounds on the performance for the class of intersection of budgeted sets and improve significantly over the state of art performance bounds. Furthermore, our analysis shows the existence of a near-optimal affine solution satisfying a nice structural property where the scenarios are partitioned into *inexpensive* and *expensive* based on a threshold and the affine solution covers only the inexpensive components using a linear solution and remaining components using a static solution. This structure is closely related to threshold policies that are widely used in many applications, and allows us to design an alternate algorithm for computing near-optimal faster affine solutions. This structural property might be of independent interest for other applications and could provide insights to design more general policies that work well in settings where affine policies could be highly sub-optimal.

In Chapters 4 and 5, we design new polices that improve significantly over affine and static policies. In fact, while affine policies provide an optimal approximation for budgeted uncertainty sets, their performance could be bad for general uncertainty sets, most notably for sets generated by conic constraints like ellipsoids. We present piecewise policies where we divide the uncertainty set into several pieces and specify an affine or a static solution for each piece. A significant challenge in designing a practical piecewise policy is to construct good pieces of the uncertainty set. In Chapter 6, we show that in the worst-case, there is no piecewise static policy with a polynomial number of pieces that has a significantly better performance than a static policy for a class of two-stage packing problems. This is quite surprising as piecewise static policies are significantly more general than static policies but still do not give a provably better solution. This motivates us to consider piecewise policies with possibly exponentially many pieces but where the pieces are not given explicitly. In particular, in Chapter 4, we introduce a new framework where the uncertainty set is implicitly partitioned into an exponential number of pieces using a threshold point. The threshold depends only on the geometry of the uncertainty set and can be computed by solving a compact linear program. This results in a tractable piecewise affine policy that performs significantly better than affine policies for many important uncertainty sets, such as ellipsoids and norm-balls, both theoretically and numerically. However, the theoretical bounds are still significantly higher than the hardness lower bounds. In Chapter 5, we significantly improve over the previous bounds by optimizing the threshold point based on both the geometry of the uncertainty set and the instance. We use insights from our previous analysis of affine policies and design a class of extended affine policies that can be compute by using linear decision rules in a lifted space. We show that that they improve significantly over previous policies for some important class of uncertainty sets. We also analyze the structure of optimal solution and show that they are closely related to threshold policies.

While this thesis has focused on two-stage robust optimization problems, it is an interesting and open question to analyze theoretically the performance of all these policies, i.e., static, affine, piecewise affine, extended affine and threshold policies or design new policies for multi-stage robust optimization problems.

References

[1] G. Dantzig, “Linear programming under uncertainty,” Management Science, vol. 1, pp. 197–206, 1955.

[2] P. Kall and S. Wallace, Stochastic programming. Wiley New York, 1994.

[3] A. Prékopa, Stochastic programming. Kluwer Academic Publishers, Dordrecht, Boston, 1995.

[4] A. Shapiro, “Stochastic programming approach to optimization under uncertainty,” Mathematical Programming, Series B, vol. 112, no. 1, pp. 183–220, 2008.

[5] A. Shapiro, D. Dentcheva, et al., Lectures on stochastic programming: modeling and theory. SIAM, 2014, vol. 16.

[6] J. Birge and F. Louveaux, Introduction to Stochastic Programming. Berlin: Springer, 1997.

[7] A. Soyster, “Convex programming with set-inclusive constraints and applications to inexact linear programming,” Operations research, vol. 21, no. 5, pp. 1154–1157, 1973.

[8] A. Ben-Tal and A. Nemirovski, “Robust convex optimization,” Mathematics of Operations Research, vol. 23, no. 4, pp. 769–805, 1998.

[9] ——, “Robust solutions of uncertain linear programs,” Operations Research Letters, vol. 25, no. 1, pp. 1–14, 1999.

[10] ——, “Robust optimization–methodology and applications,” Mathematical Programming, vol. 92, no. 3, pp. 453–480, 2002.

[11] L. El Ghaoui and H. Lebret, “Robust solutions to least-squares problems with uncertain data,” SIAM Journal on Matrix Analysis and Applications, vol. 18, pp. 1035–1064, 1997.

[12] D. Bertsimas and M. Sim, “Robust Discrete Optimization and Network Flows,” Mathematical Programming Series B, vol. 98, pp. 49–71, 2003.

[13] ——, “The price of robustness,” Operations Research, vol. 52, no. 2, pp. 35–53, 2004.

[14] D. Goldfarb and G. Iyengar, “Robust portfolio selection problems,” Mathematics of Operations Research, vol. 28, no. 1, pp. 1–38, 2003.

[15] D. Bertsimas, D. Brown, and C. Caramanis, “Theory and applications of robust optimization,” SIAM review, vol. 53, no. 3, pp. 464–501, 2011.

[16] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, Robust optimization. Princeton University press, 2009.

[17] D. Bertsimas and F. J. de Ruiter, “Duality in two-stage adaptive linear optimization: faster computation and stronger bounds,” INFORMS Journal on Computing, vol. 28, no. 3, pp. 500–511, 2016.

[18] D. Bertsimas and V. Goyal, “On the power and limitations of affine policies in two-stage adaptive Optimization,” Mathematical Programming, vol. 134, no. 2, pp. 491–531, 2012.

[19] K. Dhamdhere, V. Goyal, R Ravi, and M. Singh, “How to pay, come what may: approximation algorithms for demand-robust covering problems,” in Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on, IEEE, 2005, pp. 367–376.

[20] O. El Housni and V. Goyal, “Beyond worst-case: a probabilistic analysis of affine policies in dynamic optimization,” in Advances in Neural Information Processing Systems, 2017, pp. 4759–4767.

[21] A. Gupta, V. Nagarajan, and R Ravi, “Thresholded covering algorithms for robust and max–min optimization,” Mathematical Programming, vol. 146, no. 1-2, pp. 583–615, 2014.

[22] G. Xu and S. Burer, “A copositive approach for two-stage adjustable robust optimization with uncertain right-hand sides,” Computational Optimization and Applications, vol. 70, no. 1, pp. 33–59, 2018.

[23] J. Zhen, D. Den Hertog, and M. Sim, “Adjustable robust optimization via fourier–motzkin elimination,” Operations Research, vol. 66, no. 4, pp. 1086–1100, 2018.

[24] U. Feige, K. Jain, M. Mahdian, and V. Mirrokni, “Robust combinatorial optimization with exponential scenarios,” Lecture Notes in Computer Science, vol. 4513, pp. 439–453, 2007.

[25] D. Bertsimas, V. Goyal, and X. Sun, “A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization,” Mathematics of Operations Research, vol. 36, no. 1, pp. 24–54, 2011.

[26] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski, “Adjustable robust solutions of uncertain linear programs,” Mathematical Programming, vol. 99, no. 2, pp. 351–376, 2004.

[27] D. Bertsimas, D. Iancu, and P. Parrilo, “Optimality of affine policies in multi-stage robust optimization,” Mathematics of Operations Research, vol. 35, pp. 363–394, 2010.

[28] D. Iancu, M. Sharma, and M. Sviridenko, "Supermodularity and affine policies in dynamic robust optimization," Operations Research, vol. 61, no. 4, pp. 941–956, 2013.

[29] D. Bertsimas and A. Georghiou, "Design of near optimal decision rules in multistage adaptive mixed-integer optimization," Operations Research, vol. 63, no. 3, pp. 610–627, 2015.

[30] K. Postek and D. d. Hertog, "Multistage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set," INFORMS Journal on Computing, vol. 28, no. 3, pp. 553–574, 2016.

[31] G. A. Hanasusanto, D. Kuhn, and W. Wiesemann, "K-adaptability in two-stage robust binary programming," Operations Research, vol. 63, no. 4, pp. 877–891, 2015.

[32] O. El Housni and V. Goyal, "Piecewise static policies for two-stage adjustable robust linear optimization," Mathematical Programming, vol. 169, no. 2, pp. 649–665, 2018.

[33] X. Chen, M. Sim, P. Sun, and J. Zhang, "A linear decision-based approximation approach to stochastic programming," Operations Research, vol. 56, no. 2, pp. 344–357, 2008.

[34] J. Zhen, D. Den Hertog, and M. Sim, "Adjustable robust optimization via fourier–motzkin elimination," Operations Research, vol. 66, no. 4, pp. 1086–1100, 2018.

[35] İ. Yanıkoğlu, B. L. Gorissen, and D. den Hertog, "A survey of adjustable robust optimization," European Journal of Operational Research, vol. 277, no. 3, pp. 799–813, 2019.

[36] D. Bertsimas and H. Bidkhori, "On the performance of affine policies for two-stage adaptive optimization: a geometric perspective," Mathematical Programming, pp. 1–18, 2014.

[37] C. Bandi and D. Bertsimas, "Tractable stochastic analysis in high dimensions via robust optimization," Mathematical programming, vol. 134, no. 1, pp. 23–70, 2012.

[38] C. E. Gounaris, P. P. Repoussis, C. D. Tarantilis, W. Wiesemann, and C. A. Floudas, "An adaptive memory programming framework for the robust capacitated vehicle routing problem," Transportation Science, vol. 50, no. 4, pp. 1239–1260, 2014.

[39] D. Bertsimas and C. Caramanis, "Finite adaptability in multistage linear optimization," Automatic Control, IEEE Transactions on, vol. 55, no. 12, pp. 2751–2766, 2010.

[40] D. Bertsimas and H. Bidkhori, "On the performance of affine policies for two-stage adaptive optimization: a geometric perspective," Mathematical Programming, vol. 153, no. 2, pp. 577–594, 2015.

[41] A. Ben-Tal, O. El Housni, and V. Goyal, "A tractable approach for designing piecewise affine policies in two-stage adjustable robust optimization," Mathematical Programming, 2019.

[42] D. Bienstock and N. ÖZbay, "Computing robust basestock levels," Discrete Optimization, vol. 5, no. 2, pp. 389–414, 2008.

[43] D. Bertsimas, V. Goyal, and B. Lu, "A tight characterization of the performance of static solutions in two-stage adjustable robust linear optimization," Mathematical Programming, pp. 1–39, 2014.

[44] W. Wiesemann, D. Kuhn, and B. Rustem, "Robust resource allocations in temporal networks," Mathematical programming, vol. 135, no. 1-2, pp. 437–471, 2012.

[45] B. Dean, M. Goemans, and J. Vondrák, "Adaptivity and approximation for stochastic packing problems," in Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 2005, pp. 395–404.

[46] A. Goel and P. Indyk, "Stochastic load balancing and related problems," in Foundations of Computer Science, 1999. 40th Annual Symposium on, IEEE, 1999, pp. 579–586.

[47] V. Goyal and R. Ravi, "A ptas for the chance-constrained knapsack problem with random item sizes," Operations Research Letters, vol. 38, no. 3, pp. 161–164, 2010.

[48] P. Awasthi, V. Goyal, and B. Lu, "On the adaptivity gap in two-stage robust linear optimization under uncertain constraints," 2014.

[49] W. Hoeffding, "Probability inequalities for sums of bounded random variables," Journal of the American statistical association, vol. 58, no. 301, pp. 13–30, 1963.

[50] R. Deville, V. Fonf, and P. Hájek, "Analytic and polyhedral approximation of convex bodies in separable polyhedral banach spaces," Israel Journal of Mathematics, vol. 105, no. 1, pp. 139–154, 1998.

[51] A. Gupta, V. Nagarajan, and R Ravi, "Robust and maxmin optimization under matroid and knapsack uncertainty sets," ACM Transactions on Algorithms (TALG), vol. 12, no. 1, p. 10, 2016.

[52] N. Buchbinder, J. S. Naor, et al., "The design of competitive online algorithms via a primal–dual approach," Foundations and Trends® in Theoretical Computer Science, vol. 3, no. 2–3, pp. 93–263, 2009.

[53] N. Buchbinder and J. Naor, "Online primal-dual algorithms for covering and packing," Mathematics of Operations Research, vol. 34, no. 2, pp. 270–286, 2009.

[54] O. E. Housni and V. Goyal, "On the optimality of affine policies for budgeted uncertainty sets," arXiv preprint arXiv:1807.00163, 2018.

[55] P. Raghavan, "Probabilistic construction of deterministic algorithms: approximating packing integer programs," in Foundations of Computer Science, 1986., 27th Annual Symposium on, IEEE, 1986, pp. 10–18.

[56] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," The Annals of Mathematical Statistics, pp. 493–507, 1952.

Appendix A: Two-stage robust optimization

Lemma A.0.2. *The affine problem* (1.2) *can be formulated as the following LP*

$$
\begin{aligned}
z_{\mathsf{Aff}}(\mathcal{U}) = \min\ & \boldsymbol{c}^T \boldsymbol{x} + z \\
& z - \boldsymbol{d}^T \boldsymbol{q} \geq \boldsymbol{r}^T \boldsymbol{v} \\
& \boldsymbol{R}^T \boldsymbol{v} \geq \boldsymbol{P}^T \boldsymbol{d} \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{q} \geq \boldsymbol{V}^T \boldsymbol{r} \\
& \boldsymbol{R}^T \boldsymbol{V} \geq \boldsymbol{I}_m - \boldsymbol{B}\boldsymbol{P} \\
& \boldsymbol{q} \geq \boldsymbol{U}^T \boldsymbol{r} \\
& \boldsymbol{U}^T \boldsymbol{R} + \boldsymbol{P} \geq \boldsymbol{0} \\
& \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{v} \in \mathbb{R}_+^L,\ \boldsymbol{U} \in \mathbb{R}_+^{L\times n},\ \boldsymbol{V} \in \mathbb{R}_+^{L\times m} \\
& \boldsymbol{P} \in \mathbb{R}^{n\times m},\ \boldsymbol{q} \in \mathbb{R}^n,\ z \in \mathbb{R}.
\end{aligned}
\tag{A.1}
$$

Proof. The affine problem (1.2) has the following epigraph formulation

$$
\begin{aligned}
z_{\mathsf{Aff}}(\mathcal{U}) = \min\ & \boldsymbol{c}^T \boldsymbol{x} + z \\
& z \geq \boldsymbol{d}^T (\boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}), \quad \forall \boldsymbol{h} \in \mathcal{U} \\
& \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}) \ \geq\ \boldsymbol{h}, \quad \forall \boldsymbol{h} \in \mathcal{U} \\
& \boldsymbol{P}\boldsymbol{h} + \boldsymbol{q} \ \geq\ \boldsymbol{0}, \quad \forall \boldsymbol{h} \in \mathcal{U} \\
& \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{P} \in \mathbb{R}^{n\times m},\ \boldsymbol{q} \in \mathbb{R}^n,\ z \in \mathbb{R}.
\end{aligned}
$$

We use standard duality techniques to derive formulation (A.1). The first constraint is equivalent to

$$z - \boldsymbol{d}^T \boldsymbol{q} \geq \max_{\substack{\boldsymbol{Rh} \leq \boldsymbol{r} \\ \boldsymbol{h} \geq \boldsymbol{0}}} \boldsymbol{d}^T \boldsymbol{Ph}.$$

By taking the dual of the maximization problem, the constraint is equivalent to

$$z - \boldsymbol{d}^T \boldsymbol{q} \geq \min_{\substack{\boldsymbol{R}^T \boldsymbol{v} \geq \boldsymbol{P}^T \boldsymbol{d} \\ \boldsymbol{v} \geq \boldsymbol{0}}} \boldsymbol{r}^T \boldsymbol{v}.$$

We can then drop the min and introduce $\boldsymbol{v}$ as a variable, hence we obtain the following linear constraints

$$z - \boldsymbol{d}^T \boldsymbol{q} \geq \boldsymbol{r}^T \boldsymbol{v}$$

$$\boldsymbol{R}^T \boldsymbol{v} \geq \boldsymbol{P}^T \boldsymbol{d}$$

$$\boldsymbol{v} \in \mathbb{R}_+^L.$$

We use the same technique for the second sets of constraints, i.e.,

$$\boldsymbol{Ax} + \boldsymbol{Bq} \quad \geq \quad \max_{\substack{\boldsymbol{Rh} \leq \boldsymbol{r} \\ \boldsymbol{h} \geq \boldsymbol{0}}} \boldsymbol{h}(\boldsymbol{I}_m - \boldsymbol{BP}).$$

By taking the dual of the maximization problem for each row and dropping the min we get the following compact formulation of these constraints

$$\boldsymbol{Ax} + \boldsymbol{Bq} \geq \boldsymbol{V}^T \boldsymbol{r}$$

$$\boldsymbol{R}^T \boldsymbol{V} \geq \boldsymbol{I}_m - \boldsymbol{BP}$$

$$\boldsymbol{V} \in \mathbb{R}_+^{L \times m}.$$

Similarly, the last constraint

$$\boldsymbol{q} \quad \geq \quad \max_{\substack{\boldsymbol{Rh} \leq \boldsymbol{r} \\ \boldsymbol{h} \geq \boldsymbol{0}}} \; -\boldsymbol{Ph},$$

is equivalent to

$$\boldsymbol{q} \geq \boldsymbol{U}^T \boldsymbol{r}$$

$$\boldsymbol{U}^T \boldsymbol{R} + \boldsymbol{P} \geq \boldsymbol{0}$$

$$\boldsymbol{U} \in \mathbb{R}_+^{L \times n}.$$

Putting all together, we get the formulation (A.1).

□

Appendix B: Beyond worst-case: a probabilistic analysis of affine policies

B.1 Proofs of preliminaries

Proof of Lemma 2.2.4 We have

$$
\begin{aligned}
z_{\mathsf{AR}}(\boldsymbol{B}) &= \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \min_{\substack{\boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{h}-\boldsymbol{A}\boldsymbol{x}\\ \boldsymbol{y}\geq\boldsymbol{0}}} \boldsymbol{d}^T\boldsymbol{y}\\
&= \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}} \max_{\substack{\boldsymbol{B}^T\boldsymbol{w}\leq\boldsymbol{d}\\ \boldsymbol{w}\geq\boldsymbol{0}}} (\boldsymbol{h}-\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{w} \qquad \text{(B.1)}\\
&= \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{w}\in\mathcal{W}} -(\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{w} + \max_{\substack{\boldsymbol{R}\boldsymbol{h}\leq\boldsymbol{r}\\ \boldsymbol{h}\geq\boldsymbol{0}}} \boldsymbol{h}^T\boldsymbol{w}\\
&= \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{w}\in\mathcal{W}} -(\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{w} + \min_{\substack{\boldsymbol{R}^T\boldsymbol{\lambda}\geq\boldsymbol{w}\\ \boldsymbol{\lambda}\geq\boldsymbol{0}}} \boldsymbol{r}^T\boldsymbol{\lambda}\\
&= z_{\mathsf{d-AR}}(\boldsymbol{B}).
\end{aligned}
$$

where the second equality holds by taking the dual of the inner minimization problem, the third equality follows from switching the two max, and the fourth one by taking the dual of the second maximization problem.

Proof of Lemma 2.2.6 We restate the same proof in [18] in our setting. First, since the adjustable problem is a relaxation of the affine problem then $z_{\mathsf{d-AR}}(\boldsymbol{B}) \leq z_{\mathsf{d-Aff}}(\boldsymbol{B})$.

Now let us prove the other inequality. Consider $\mathcal{W} = \{\boldsymbol{w} \in \mathbb{R}^m_+ \mid \boldsymbol{B}^T\boldsymbol{w} \leq \boldsymbol{d}\}$ which is a simplex. Note that $\boldsymbol{0}$ is always an extreme point of the simplex $\mathcal{W}$ and denote $\boldsymbol{w}^1, \boldsymbol{w}^2, \ldots, \boldsymbol{w}^m$ the remaining m points. In particular, we have for any $\boldsymbol{w} \in \mathcal{W}$

$$\boldsymbol{w} = \sum_{j=1}^{m} \alpha_j \boldsymbol{w}^j = \boldsymbol{Q}\alpha$$

where $\sum_{j=1}^{m} \alpha_j \leq 1$ and $\boldsymbol{Q} = \left[\boldsymbol{w}^1 | \boldsymbol{w}^2 | \ldots | \boldsymbol{w}^m\right]$. Note that $\boldsymbol{Q}$ is invertible since $\boldsymbol{w}^1, \boldsymbol{w}^2, \ldots, \boldsymbol{w}^m$ are linearly independent. Hence, $\boldsymbol{\alpha} = \boldsymbol{Q}^{-1}\boldsymbol{w}$. Denote $\boldsymbol{x}^*, \boldsymbol{\lambda}^*(\boldsymbol{w}), \boldsymbol{w} \in \mathcal{W}$, an optimal solution of the adjustable problem (2.2). We define the following affine solution $\boldsymbol{x} = \boldsymbol{x}^*$ and for $\boldsymbol{w} \in \mathcal{W}$,

$$\boldsymbol{\lambda}(\boldsymbol{w}) = \boldsymbol{P}\boldsymbol{Q}^{-1}\boldsymbol{w}$$

where

$$\boldsymbol{P} = \left[\boldsymbol{\lambda}^*(\boldsymbol{w}^1) | \boldsymbol{\lambda}^*(\boldsymbol{w}^2) | \ldots | \boldsymbol{\lambda}^*(\boldsymbol{w}^m)\right].$$

In particular, we have

$$\boldsymbol{\lambda}(\boldsymbol{w}) = \sum_{j=1}^{m} \alpha_j \boldsymbol{\lambda}^*(\boldsymbol{w}^j).$$

Let us first check the feasbility of the solution. We have,

$$\boldsymbol{R}^T \boldsymbol{\lambda}(\boldsymbol{w}) = \sum_{j=1}^{m} \alpha_j \boldsymbol{R}^T \boldsymbol{\lambda}^*(\boldsymbol{w}^j) \geq \sum_{j=1}^{m} \alpha_j \boldsymbol{w}^j = \boldsymbol{w}$$

where the inequality follows from the feasibility of the adjustable solution. Therefore,

$$\begin{aligned}
z_{\mathsf{d-Aff}}(\boldsymbol{B}) &\leq \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{w}\in\mathcal{W}} \ (-\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{w} + \boldsymbol{r}^T\boldsymbol{\lambda}(\boldsymbol{w}) \\
&= \boldsymbol{c}^T\boldsymbol{x}^* + \max_{\boldsymbol{\alpha}} \ (-\boldsymbol{A}\boldsymbol{x}^*)^T\boldsymbol{w} + \sum_{j=1}^{m}\alpha_j\boldsymbol{r}^T\boldsymbol{\lambda}^*(\boldsymbol{w}^j) \\
&= \boldsymbol{c}^T\boldsymbol{x}^* + \max_{\boldsymbol{\alpha}} \ \sum_{j=1}^{m}\alpha_j\left((-\boldsymbol{A}\boldsymbol{x}^*)^T\boldsymbol{w}^j + \boldsymbol{r}^T\boldsymbol{\lambda}^*(\boldsymbol{w}^j)\right) \\
&\leq \boldsymbol{c}^T\boldsymbol{x}^* + \max_{\boldsymbol{w}\in\mathcal{W}}\left((-\boldsymbol{A}\boldsymbol{x}^*)^T\boldsymbol{w} + \boldsymbol{r}^T\boldsymbol{\lambda}^*(\boldsymbol{w})\right)\max_{\boldsymbol{\alpha}} \ \sum_{j=1}^{m}\alpha_j \leq z_{\mathsf{d-AR}}(\boldsymbol{B})
\end{aligned}$$

where the last inequality holds because $\sum_{j=1}^{m}\alpha_j \leq 1$. We conclude that $z_{\mathsf{d-Aff}}(\boldsymbol{B}) = z_{\mathsf{d-AR}}(\boldsymbol{B})$.

Proof of Lemma 2.2.7 First the inequality $z_{\mathsf{d-AR}}(\boldsymbol{B}) \leq z_{\mathsf{d-Aff}}(\boldsymbol{B})$ is straightforward since the adjustable problem(1.1) is a relaxation of the affine problem (1.2). On the other hand, since $\mathcal{W} \subseteq \kappa \cdot \mathcal{S}$ then,

$$z_{\mathsf{d-Aff}}(\boldsymbol{B}) \leq \kappa \cdot z_{\mathsf{d-Aff}}(\boldsymbol{B},\mathcal{S})$$

where we denote $z_{\mathsf{d-Aff}}(\boldsymbol{B},\mathcal{S})$ the dualized affine problem over $\mathcal{S}$ (it's the same problem as $z_{\mathsf{d-Aff}}(\boldsymbol{B})$ where we only replace $\mathcal{W}$ by $\mathcal{S}$). Since $\mathcal{S}$ is a simplex, from Lemma 2.2.6, we have $z_{\mathsf{d-Aff}}(\boldsymbol{B},\mathcal{S}) = z_{\mathsf{d-AR}}(\boldsymbol{B},\mathcal{S})$. Moreover, $z_{\mathsf{d-AR}}(\boldsymbol{B},\mathcal{S}) \leq z_{\mathsf{d-AR}}(\boldsymbol{B})$ because $\mathcal{S} \subseteq \mathcal{W}$. We conclude that

$$z_{\mathsf{d-AR}}(\boldsymbol{B}) \leq z_{\mathsf{d-Aff}}(\boldsymbol{B}) \leq \kappa \cdot z_{\mathsf{d-AR}}(\boldsymbol{B}).$$

Furthermore, since $z_{\mathsf{d-AR}}(\boldsymbol{B}) = z_{\mathsf{AR}}(\boldsymbol{B})$ from Lemma 2.2.4 and $z_{\mathsf{d-Aff}}(\boldsymbol{B}) = z_{\mathsf{Aff}}(\boldsymbol{B})$ from Lemma 2.2.5, then

$$z_{\mathsf{AR}}(\boldsymbol{B}) \leq z_{\mathsf{Aff}}(\boldsymbol{B}) \leq \kappa \cdot z_{\mathsf{AR}}(\boldsymbol{B}).$$

B.2 Hoeffding's inequality

Hoeffding's inequality[49]. Let $Z_1, \ldots, Z_n$ be independent bounded random variables with $Z_i \in [a, b]$ for all $i \in [n]$ and denote $Z = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Therefore,

$$\mathbb{P}(Z - \mathbb{E}(Z) \leq -\tau) \leq \exp\left(\frac{-2n\tau^2}{(b-a)^2}\right).$$

B.3 MIP formulation for the empirical section

MIP formulation for the separation adjustable problem. The separation problem (2.18) can be formulated as the following MIP

$$\begin{aligned}
\max \quad & \sum_{i=1}^{m} \sum_{j=-\Delta_{\mathcal{W}}}^{s} \sum_{k=-\Delta_{\mathcal{U}}}^{s} \frac{1}{2^{j+k}} \cdot \gamma_{ijk} - (\boldsymbol{A}\hat{\boldsymbol{x}})^T \boldsymbol{w} & & \\
& \boldsymbol{w} = \sum_{i=1}^{m} \sum_{j=-\Delta_{\mathcal{W}}}^{s} \frac{\beta_{ij}}{2^j} \cdot \boldsymbol{e}_i & & \\
& \boldsymbol{h} = \sum_{i=1}^{m} \sum_{k=-\Delta_{\mathcal{U}}}^{s} \frac{\alpha_{ik}}{2^k} \cdot \boldsymbol{e}_i & & \\
& \gamma_{ijk} \leq \beta_{ij} & \forall i \in [m], j \in [-\Delta_{\mathcal{U}}, s], k \in [-\Delta_{\mathcal{W}}, s] & \quad \text{(B.2)} \\
& \gamma_{ijk} \leq \alpha_{ik} & \forall i \in [m], j \in [-\Delta_{\mathcal{U}}, s], k \in [-\Delta_{\mathcal{W}}, s] & \\
& \gamma_{ijk} + 1 \geq \alpha_{ik} + \beta_{ij} & \forall i \in [m], j \in [-\Delta_{\mathcal{U}}, s], k \in [-\Delta_{\mathcal{W}}, s] & \\
& \alpha_{ik}, \beta_{ik}, \gamma_{ijk} \in \{0, 1\} & \forall i \in [m], j \in [-\Delta_{\mathcal{U}}, s], k \in [-\Delta_{\mathcal{W}}, s] & \\
& \boldsymbol{R}\boldsymbol{h} \leq \boldsymbol{r} & & \\
& \boldsymbol{B}^T \boldsymbol{w} \leq \boldsymbol{d} & &
\end{aligned}$$

where $s = \lceil \log_2 \left(\frac{m}{\epsilon}\right) \rceil$, $\Delta_{\mathcal{W}}$ is an upper bound on any component of $w \in \mathcal{W}$, $\Delta_{\mathcal{U}}$ is an upper bound on any component of $h \in \mathcal{U}$ and ϵ is the accuracy of the problem.

Proof. The separation problem (2.18) is equivalent to solving the following problem for given $\hat{\boldsymbol{x}}$

$$\max_{\substack{\boldsymbol{h}\in\mathcal{U}\\ \boldsymbol{w}\in\mathcal{W}}} \boldsymbol{h}^T\boldsymbol{w} - (\boldsymbol{A}\hat{\boldsymbol{x}})^T\boldsymbol{w}$$

The constraints of the above problem are linear and the second term in the objective function is linear as well. So we will focus only on the first term $\boldsymbol{h}^T\boldsymbol{w}$ which is a bilinear function and write it in terms of linear constraints and binary variables. Let us write $\boldsymbol{h} = \sum_{i=1}^m h_i\boldsymbol{e}_i$. For all $i \in [m]$ we digitize the component $\boldsymbol{h}_i$ as follows

$$h_i = \sum_{k=-\Delta_{\mathcal{U}}}^{s} \frac{\alpha_{ik}}{2^k}$$

where $s = \lceil \log_2\left(\frac{m}{\epsilon}\right)\rceil$, $\Delta_{\mathcal{U}}$ is an upper bound on any h_i and α_{ik} are binary variables. This digitization gives an approximation to h_i within $\frac{\epsilon}{m}$ which translates to an accuracy of ϵ in the objective function. We have

$$\boldsymbol{h} = \sum_{i=1}^{m}\sum_{k=-\Delta_{\mathcal{U}}}^{s} \frac{\alpha_{ik}}{2^k}\cdot\boldsymbol{e}_i.$$

Similarly, we have

$$\boldsymbol{w} = \sum_{i=1}^{m}\sum_{j=-\Delta_{\mathcal{W}}}^{s} \frac{\beta_{ij}}{2^j}\cdot\boldsymbol{e}_i$$

where $\Delta_{\mathcal{W}}$ is an upper bound on any component of $w \in \mathcal{W}$. Therefore, the first term in the objective function becomes

$$\sum_{i=1}^{m}\sum_{j=-\Delta_{\mathcal{W}}}^{s}\sum_{k=-\Delta_{\mathcal{U}}}^{s} \frac{1}{2^{j+k}}\cdot\alpha_{ik}\beta_{ij}.$$

The final step is to linearize the term $\alpha_{ik}\beta_{ij}$. We set, $\alpha_{ik}\beta_{ij} = \gamma_{ijk}$ where again γ_{ijk} is a binary variable. Since all the variables here are binary we can express γ_{ijk} using only linear constraints as follows

$$\gamma_{ijk} \le \beta_{ij}$$

$$\gamma_{ijk} \le \alpha_{ik}$$

$$\gamma_{ijk} + 1 \ge \alpha_{ik} + \beta_{ij}$$

which leads to formulation (B.2).

□

Appendix C: Affine policies for budget of uncertainty sets

C.1 Proof of Claims in Section 3.2.3

Proof of Claim 3.2.6.

$$
\begin{aligned}
z(\mathcal{W}) &= \min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \boldsymbol{B}\boldsymbol{y} \geq \sum_{i \in \mathcal{W}} \boldsymbol{e}_i \right\} \\
&= \min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \sum_{j=1}^{n} B_{ij} y_j \geq 1, \ \forall i \in \mathcal{W} \right\} \\
&= \min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \sum_{j=1}^{n} \frac{w_i B_{ij}}{\max_{k \in \mathcal{W}} (w_k B_{kj})} \cdot \max_{k \in \mathcal{W}} (w_k B_{kj}) \cdot y_j \geq w_i, \ \forall i \in \mathcal{W} \right\} \\
&= \min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \boldsymbol{d}^T \boldsymbol{y} \;\middle|\; \sum_{j=1}^{n} \hat{B}_{ij} \cdot \max_{k \in \mathcal{W}} (w_k B_{kj}) \cdot y_j \geq w_i, \ \forall i \in \mathcal{W} \right\} \\
&= \min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \sum_{j=1}^{n} \frac{d_j}{\max_{k \in \mathcal{W}} (w_k B_{kj})} \cdot y_j \;\middle|\; \sum_{j=1}^{n} \hat{B}_{ij} y_j \geq w_i, \ \forall i \in \mathcal{W} \right\} \\
&= \min_{\boldsymbol{y} \geq \boldsymbol{0}} \left\{ \eta\gamma \sum_{j=1}^{n} \hat{d}_j y_j \;\middle|\; \sum_{j=1}^{n} \hat{B}_{ij} y_j \geq w_i, \ \forall i \in \mathcal{W} \right\} \\
&= \eta\gamma \cdot \hat{z}(\mathcal{W}).
\end{aligned}
$$

Proof of Claim 3.2.7. Consider $j \in [n]$, by the feasibility of the solution $\boldsymbol{z}^*$ we have,

$$
\sum_{i \in \mathcal{J}} \hat{B}_{ij} \cdot z_i^* \leq \hat{d}_j,
$$

and consequently

$$\sum_{i\in\mathcal{J}} \hat{B}_{ij}\frac{2\lfloor z_i^*\rfloor}{\eta} \leq \sum_{i\in\mathcal{J}} \hat{B}_{ij}\frac{2z_i^*}{\eta} \leq \frac{2\hat{d}_j}{\eta}.$$

Hence,

$$\begin{aligned}
\mathbb{P}\left(\sum_{i\in\mathcal{J}} \hat{B}_{ij}\cdot\frac{2Z_i}{\eta} > \hat{d}_j\right) &= \mathbb{P}\left(\sum_{i\in\mathcal{J}} \hat{B}_{ij}\cdot\frac{2(\lfloor z_i^*\rfloor + \xi_i)}{\eta} > \hat{d}_j\right) \\
&\leq \mathbb{P}\left(\sum_{i\in\mathcal{J}} \hat{B}_{ij}\frac{2\xi_i}{\eta} > (1-\frac{2}{\eta})\hat{d}_j\right) \\
&= \mathbb{P}\left(\sum_{i\in\mathcal{J}} \hat{B}_{ij}\xi_i > (\frac{\eta}{2}-1)\hat{d}_j\right).
\end{aligned}$$

Now, we apply the Chernoff inequality in Lemma C.2.1 with $\delta = \frac{\eta}{2} - 2$ and $\Xi = \sum_{i\in\mathcal{J}} \hat{B}_{ij}\xi_i$. Note that $\delta = 2\cdot\frac{\log n}{\log\log n} - 2 > 0$ for sufficiently large n. Moreover, we have for all $i\in\mathcal{J}, j\in\mathcal{J}$, $\hat{B}_{ij}\in[0,1]$ and

$$\mathbb{E}\left(\sum_{i\in\mathcal{J}} \hat{B}_{ij}\xi_i\right) = \sum_{i\in\mathcal{J}} \hat{B}_{ij}(z_i^* - \lfloor z_i^*\rfloor) \leq \sum_{i\in\mathcal{J}} \hat{B}_{ij}z_i^* \leq \hat{d}_j.$$

Therefore the Chernoff bound gives,

$$\mathbb{P}\left(\sum_{i\in\mathcal{J}} \hat{B}_{ij}\xi_i > (\frac{\eta}{2}-1)\hat{d}_j\right) \leq \left(\frac{e^{\frac{\eta}{2}-1}}{(\frac{\eta}{2}-2)^{\frac{\eta}{2}-2}}\right)^{\hat{d}_j}.$$

Recall $\eta = 4\frac{\log n}{\log\log n}$. Hence the RHS is equivalent to

$$\begin{aligned}
\left(\frac{e^{\frac{\eta}{2}-1}}{(\frac{\eta}{2}-2)^{\frac{\eta}{2}-2}}\right)^{\hat{d}_j} = O\left(\frac{e^{\frac{\eta\hat{d}_j}{2}}}{(\frac{\eta}{2})^{\frac{\hat{d}_j\eta}{2}}}\right) &= O\left(\exp\left(\hat{d}_j\frac{\eta}{2}\left(1-\log(\frac{\eta}{2})\right)\right)\right) \\
&= O\left(\exp\left(-\hat{d}_j\frac{\eta}{2}\log\eta\right)\right) \\
&= O\left(\exp\left(-2\hat{d}_j\frac{\log n}{\log\log n}\log\left(4\frac{\log n}{\log\log n}\right)\right)\right) \\
&= O\left(\exp\left(-2\hat{d}_j\log n\right)\right) = O\left(\frac{1}{n^{2\hat{d}_j}}\right) \leq \frac{c}{n^2},
\end{aligned}$$

for some constant c. The last inequality holds because $\hat{d}_j \geq 1$. Hence,

$$\mathbb{P}\left(\sum_{i\in\mathcal{J}} \hat{B}_{ij} \cdot \frac{2Z_i}{\eta} > \hat{d}_j\right) \leq \frac{c}{n^2}.$$

Therefore by a union bound we have,

$$\mathbb{P}\left(\sum_{i\in\mathcal{J}} \hat{B}_{ij} \frac{2Z_i}{\eta} > \hat{d}_j,\ \exists j \in [n]\right) \leq \sum_{i=1}^{n} \mathbb{P}\left(\sum_{i\in\mathcal{J}} \hat{B}_{ij} \frac{2Z_i}{\eta} > \hat{d}_j\right) \leq \frac{c}{n}.$$

Therefore,

$$\mathbb{P}\left(\sum_{i\in\mathcal{J}} \hat{B}_{ij} \frac{2Z_i}{\eta} \leq \hat{d}_j,\ \forall j \in [n]\right) \geq 1 - \frac{c}{n} = 1 - O(\frac{1}{n}).$$

Proof of Claim 3.2.8. We have,

$$\sum_{i\in\mathcal{J}} w_i Z_i = \lambda + \sum_{i\in\mathcal{J}} w_i \xi_i,$$

where

$$\lambda = \sum_{i\in\mathcal{J}} w_i \lfloor z_i^* \rfloor.$$

We apply the Chernoff inequality in Lemma C.2.2 with $\delta = \frac{1}{2}$ and $\Xi = \lambda + \sum_{i\in\mathcal{J}} w_i \xi_i$. Note that

$$\mathbb{E}(\Xi) = \sum_{i\in\mathcal{J}} w_i z_i^* > 1.$$

Hence,

$$\begin{aligned}
\mathbb{P}\left(\sum_{i\in\mathcal{J}} w_i Z_i > \frac{1}{2}\right) &= \mathbb{P}\left(\Xi > \frac{1}{2}\right) \\
&\geq \mathbb{P}\left(\Xi > \frac{\mathbb{E}(\Xi)}{2}\right) \\
&\geq 1 - e^{-\frac{\mathbb{E}(\Xi)}{8}} \geq 1 - e^{-\frac{1}{8}}.
\end{aligned}$$

C.2 Chernoff bounds

Lemma C.2.1 (Chernoff Bound 1). *Let $\xi_1, \xi_2, \dots, \xi_r$ be independent Bernoulli trials. Denote $\Xi = \sum_{i=1}^{r} \alpha_i \xi_i$ where $\alpha_1, \dots, \alpha_r$ are reals in $[0,1]$. Let $s > 0$ such that $\mathbb{E}(\Xi) \leq s$. Then for any $\delta > 0$,*

$$\mathbb{P}(\Xi \geq (1+\delta)s) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{s}.$$

The Chernoff bound in Lemma C.2.1 is a slight variant of the Raghavan-Spencer inequality (Theorem 1 in [55]). The proof is along the same lines as in [55]. For completeness, we are providing it below.

Proof. From Markov's inequality we have for all $t > 0$,

$$\mathbb{P}(\Xi \geq (1+\delta)s) = \mathbb{P}(e^{t\Xi} \geq e^{t(1+\delta)s}) \leq \frac{\mathbb{E}(e^{t\Xi})}{e^{t(1+\delta)s}}.$$

Denote p_i the parameter of the Bernoulli ξ_i. By independence, we have

$$\mathbb{E}(e^{t\Xi}) = \prod_{i=1}^{r} \mathbb{E}(e^{t\alpha_i \xi_i}) = \prod_{i=1}^{r} \left(p_i e^{t\alpha_i} + 1 - p_i\right) \leq \prod_{i=1}^{r} \exp\left(p_i(e^{t\alpha_i} - 1)\right)$$

where the inequality holds because $1 + x \leq e^{x}$ for all $x \in \mathbb{R}$. By taking $t = \ln(1+\delta) > 0$, the right hand side becomes

$$\begin{aligned}\prod_{i=1}^{r} \exp\left(p_i((1+\delta)^{\alpha_i} - 1)\right) &\leq \prod_{i=1}^{r} \exp\left(p_i \delta \alpha_i\right) \\ &= \exp\left(\delta \cdot \mathbb{E}(\Xi)\right) \leq e^{\delta s},\end{aligned}$$

where the first inequality holds because $(1+x)^{\alpha} \leq 1 + \alpha x$ for any $x \geq 0$ and $\alpha \in [0,1]$ and the second one because $s \geq \mathbb{E}(\Xi) = \sum_{i=1}^{r} \alpha_i p_i$. Hence, we have

$$\mathbb{E}(e^{t\Xi}) \leq e^{\delta s}.$$

On the other hand,

$$e^{t(1+\delta)s} = (1+\delta)^{(1+\delta)s}.$$

Therefore,

$$\mathbb{P}(\Xi \geq (1+\delta)s) \leq \left(\frac{e^{\delta s}}{(1+\delta)^{(1+\delta)s}}\right) = \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{s}.$$

□

Lemma C.2.2 (Chernoff Bound 2). *Let $\xi_1, \xi_2, \ldots, \xi_r$ be independent Bernoulli trials. Denote*

$$\Xi = \lambda + \sum_{i=1}^{r} \alpha_i \xi_i$$

where $\lambda \in \mathbb{R}_+$ and $\alpha_1, \ldots, \alpha_r$ are reals in $(0,1]$. Denote $\mu = \mathbb{E}(\Xi)$. Then for any $0 < \delta < 1$,

$$\mathbb{P}(\Xi > (1-\delta)\mu) > 1 - e^{\frac{-\delta^2 \mu}{2}}.$$

This is a slight variant of the lower tail Chernoff bound [56]. The proof is along the same lines as in [56]. For completeness, we are providing it below.

Proof. We show equivalently that

$$\mathbb{P}(\Xi \leq (1-\delta)\mu) \leq e^{\frac{-\delta^2 \mu}{2}}.$$

From Markov's inequality we have for all $t < 0$,

$$\mathbb{P}(\Xi \leq (1-\delta)\mu) = \mathbb{P}(e^{t\Xi} \geq e^{t(1-\delta)\mu}) \leq \frac{\mathbb{E}(e^{t\Xi})}{e^{t(1-\delta)\mu}}.$$

Denote p_i the parameter of the Bernoulli ξ_i. By independence, we have

$$\mathbb{E}(e^{t\Xi}) = e^{t\lambda} \prod_{i=1}^{r} \mathbb{E}(e^{t\alpha_i \xi_i}) = e^{t\lambda} \prod_{i=1}^{r} \left(p_i e^{t\alpha_i} + 1 - p_i\right) \leq e^{t\lambda} \prod_{i=1}^{r} \exp\left(p_i(e^{t\alpha_i} - 1)\right),$$

where the inequality holds because $1 + x \leq e^x$ for all $x \in \mathbb{R}$. We take $t = \ln(1-\delta) < 0$. We have $t \leq -\delta$, hence

$$e^{t\lambda} \leq e^{-\delta\lambda}.$$

Moreover,

$$\prod_{i=1}^{r} \exp\left(p_i(e^{t\alpha_i} - 1)\right) = \prod_{i=1}^{r} \exp\left(p_i((1-\delta)^{\alpha_i} - 1)\right) \leq \prod_{i=1}^{r} \exp\left(-p_i \delta \alpha_i\right),$$

where the inequality holds because $(1-x)^\alpha \leq 1 - \alpha x$ for any $0 < x < 1$ and $\alpha \in [0,1]$. Therefore,

$$\mathbb{E}(e^{t\Xi}) \leq e^{-\delta\lambda} \prod_{i=1}^{r} \exp\left(-p_i \delta \alpha_i\right) = e^{-\delta\mu}.$$

On the other hand,

$$e^{t(1-\delta)\mu} = (1-\delta)^{(1-\delta)\mu}.$$

Therefore,

$$\mathbb{P}(\Xi \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta\mu}}{(1-\delta)^{(1-\delta)\mu}}\right) = \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}.$$

Finally, we have for any $0 < \delta < 1$,

$$\ln(1-\delta) \geq -\delta + \frac{\delta^2}{2}$$

which implies

$$(1-\delta) \cdot \ln(1-\delta) \geq -\delta + \frac{\delta^2}{2}$$

and consequently

$$\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \leq e^{\frac{-\delta^2\mu}{2}}.$$

□

C.3 Additional numerical experiments for general uncertainty sets

The formulation (3.19) provides an approximate affine policy for solving our two-stage adjustable problem under any uncertainty set even though the theoretical performance bound for the approximate affine policy holds only for the case of single budget of uncertainty set. In this section, we test numerically the approximate affine policy given by (3.19) for general uncertainty sets both in terms of performance and running time. We consider the following uncertainty sets for our extended numerical experiments.

$$\mathcal{U}_3 = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} h_i^2 \leq 1 \right\} \tag{C.1}$$

$$\mathcal{U}_4 = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} w_{\ell i} h_i \leq 1, \forall \ell = 1,2 \right\} \tag{C.2}$$

$$\mathcal{U}_5 = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} w_{\ell i} h_i \leq 1, \forall \ell = 1,\ldots,5 \right\}, \tag{C.3}$$

where $\mathcal{U}_3$ is the unit hypersphere uncertainty set, $\mathcal{U}_4$ and $\mathcal{U}_5$ are respectively intersection of two and five budget of uncertainty sets. For $\mathcal{U}_4$ and $\mathcal{U}_5$, we choose $\boldsymbol{w}_\ell$ to be normalized Gaussian vectors, i.e., $w_{\ell i} = |G_i|/\|\boldsymbol{G}\|_2$ where G_i are i.i.d. standard Gaussians. We use the same test instances and the same notations as in Section 3.5. We present the results of these computational experiments in Table C.1.

Results. We observe from Table C.1 that our algorithm is significantly faster than the optimal affine policy up to 100 factor of magnitude even for general uncertainty sets. The average running

m	$T_{\mathsf{aff}}(s)$	$T_{\mathsf{Alg}}(s)$	$z_{\mathsf{Alg}}/z_{\mathsf{Aff}}$
10	0.222	0.165	1.735
20	0.945	0.608	1.870
30	2.868	1.445	1.911
40.	6.653	2.533	1.952
50	15.00	4.113	1.970
60	32.34	6.148	1.987
70	69.83	9.639	2.004
80	254.1	21.59	2.010
90	500.7	30.13	2.025
100	907.6	41.09	2.030

(a) Uncertainty set (C.1)

m	$T_{\mathsf{aff}}(s)$	$T_{\mathsf{Alg}}(s)$	$z_{\mathsf{Alg}}/z_{\mathsf{Aff}}$
10	0.020	0.040	1.132
20	0.289	0.165	1.145
30	1.050	0.328	1.135
40	5.014	0.851	1.122
50	19.48	1.497	1.120
60	77.13	3.048	1.116
70	184.5	5.279	1.113
80	392.7	7.984	1.116
90	872.9	11.19	1.115
100	1199	11.83	1.109

(b) Uncertainty set (C.2)

m	$T_{\mathsf{aff}}(s)$	$T_{\mathsf{Alg}}(s)$	$z_{\mathsf{Alg}}/z_{\mathsf{Aff}}$
10	0.031	0.051	1.212
20	0.434	0.197	1.190
30	2.362	0.581	1.185
40	8.534	1.177	1.176
50	28.34	1.979	1.168
60	75.38	3.815	1.164
70	176.9	5.648	1.159
80	388.2	8.145	1.152
90	845.5	11.68	1.154
100	1133	11.48	1.147

(c) Uncertainty set (C.3)

Table C.1: Comparison on the performance and computation time of the optimal affine policy and our approximate affine policy. For 20 instances, we compute $z_{\mathsf{Alg}}(\mathcal{U})/z_{\mathsf{Aff}}(\mathcal{U})$ for the uncertainty sets (C.1), (C.2) and (C.3). Here, $T_{\mathsf{Alg}}(s)$ denotes the running time for our approximate affine policy and $T_{\mathsf{aff}}(s)$ denotes the running time for affine policy in seconds. These results are obtained using Gurobi 7.0.2 on a 16-core server with 2.93GHz processor and 56GB RAM.

time of our algorithm is few seconds and scales very well with the dimension of the problem while computing the optimal affine policy becomes challenging for large size instances. Furthermore, we observe that the gap between our affine solution and the optimal one is within 20% for intersection of budget of uncertainty sets $\mathcal{U}_4$ and $\mathcal{U}_5$ and does not increase with the dimension m. However, for the hypersphere uncertainty set $\mathcal{U}_3$, we observe that the gap between our policy and the optimal affine policy is larger as compared to other uncertainty sets and does increase in this case with the dimension m. For instance, the gap is more than a factor 2 for $m = 100$.

C.4 Proof of Lemma 3.6.1

First let us show that $z_{\mathsf{AR}} = 0$. We consider the following solution for the adjustable problem

$$x_i = \begin{cases} 1 & \text{if } i \in J_1 \\ 0 & \text{if } i \in J_2. \end{cases}$$

Consider a scenario $\boldsymbol{h}$ that is an extreme point of $\mathcal{U}$. In particular, we have $h_i \in \{0,1\}^m$ for all i and $\sum_{i=1}^m h_i = m/2$. Consider $\tilde{J}_1(\boldsymbol{h})$ and $\tilde{J}_2(\boldsymbol{h})$ respectively subsets of J_1 and J_2 in which the demand is realized. In particular,

$$\tilde{J}_1(\boldsymbol{h}) = \{i \in J_1 \mid h_i = 1\}$$

$$\tilde{J}_2(\boldsymbol{h}) = \{i \in J_2 \mid h_i = 1\}.$$

We have,

$$|\tilde{J}_1(\boldsymbol{h})| + |\tilde{J}_2(\boldsymbol{h})| = m/2.$$

Note that the first stage solution $\boldsymbol{x}$ covers the demand of the nodes in $\tilde{J}_1(\boldsymbol{h})$ because $x_i = 1$ for all $i \in J_1$. On the other side, we cover demand in $\tilde{J}_2(\boldsymbol{h})$ by sending inventory from $J_1 \setminus \tilde{J}_1(\boldsymbol{h})$ to $\tilde{J}_2(\boldsymbol{h})$ in the second stage. This is possible because

$$|J_1 \setminus \tilde{J}_1(\boldsymbol{h})| = m/2 - |\tilde{J}_1(\boldsymbol{h})| = |\tilde{J}_2(\boldsymbol{h})|.$$

The cost of sending inventory from $J_1 \setminus \tilde{J}_1(\boldsymbol{h})$ to $\tilde{J}_2(\boldsymbol{h})$ is 0 because all directed distances from J_1 to J_2 are zero. In particular, we consider a matching M from $J_1 \setminus \tilde{J}_1(\boldsymbol{h})$ to $\tilde{J}_2(\boldsymbol{h})$. We define the following second stage solution

$$y_{ij}(\boldsymbol{h}) = \begin{cases} 1 & \text{if } (i,j) \in M \\ 0 & \text{otherwise.} \end{cases}$$

We have $\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{h})$ is feasible for the adjustable problem and its corresponding cost is 0. Therefore,

$z_{\mathsf{AR}}(\mathcal{U}) = 0$.

Now, let us show that $z_{\mathsf{Aff}}(\mathcal{U}) = m/2 - 1$. Given the distances in the graph, flow can be sent only from J_1 to J_2. In particular, we can rewrite the covering constraints of the problem as follows

$$\forall j \in J_2 \qquad x_j + \sum_{i \in J_1} y_{ij}(\boldsymbol{h}) \geq h_j \tag{C.4}$$

$$\forall i \in J_1 \qquad x_i - \sum_{j \in J_2} y_{ij}(\boldsymbol{h}) \geq h_i. \tag{C.5}$$

Consider $\boldsymbol{x}$ and $\boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}$ a feasible affine solution. The number of rows of $\boldsymbol{P}$ is the number of edges in the graph. The number of columns of $\boldsymbol{P}$ is m which is the total number of nodes. In particular,

$$\forall i \in J_1, \ \forall j \in J_2 \qquad y_{ij}(\boldsymbol{h}) = \sum_{\ell=1}^{m} P_{(i,j),\ell} h_\ell + q_{ij},$$

where $P_{(i,j),\ell}$ denotes the component of $\boldsymbol{P}$ corresponding to edge (i, j) and node ℓ. Nodes of J_1 should be covered in the first stage because flow can not be sent to J_1 in the second stage. Moreover, J_1 is covered at a zero cost. In particular we have,

$$x_i = 1 \qquad \forall i \in J_1.$$

Let fix $i \in J_1$. Consider $\boldsymbol{h} \in \mathcal{U}$ such that $h_i = 1$. From (C.5), we have,

$$x_i \geq \sum_{j \in J_2} y_{ij}(\boldsymbol{h}) + 1.$$

Moreover, we know that $x_i \leq 1$ and $y_{ij}(\boldsymbol{h}) \geq 0$ for all $j \in J_2$. This implies that for all $j \in J_2$,

$$y_{ij}(\boldsymbol{h}) = 0,$$

which is equivalent to

$$P_{(i,j),i} + q_{ij} + \sum_{\ell=1,\ell\neq i}^{m} P_{(i,j),\ell} h_\ell = 0$$

for any $\boldsymbol{h} \in \mathcal{U}$ such that $h_i = 1$. Therefore, for any $i \in J_1$, for any $j \in J_2$,

$$P_{(i,j),i} + q_{ij} = 0,$$

and

$$P_{(i,j),\ell} = 0 \qquad \forall \ell \neq i. \tag{C.6}$$

Now fix $\ell \in J_1$ and $j \in J_2$. Consider the following scenario

$$\begin{cases} h_\ell = 0 \\ h_i = 1 \qquad \forall i \in J_1 \setminus \{\ell\} \\ h_j = 1 \\ h_k = 0 \qquad \forall k \in J_2 \setminus \{j\}. \end{cases}$$

From (C.4), we have for all $j \in J_2$,

$$x_j + \sum_{i \in J_1} y_{ij}(\boldsymbol{h}) \geq 1.$$

Since there is a unit demand at node $i \in J_1$ for any $i \neq \ell$, we can not send inventory from i to any node in J_2. In particular, $y_{ij}(\boldsymbol{h}) = 0$ for any $i \neq \ell$. This implies from the last inequality that

$$x_j + y_{\ell j}(\boldsymbol{h}) \geq 1,$$

which is equivalent to

$$x_j + P_{(\ell,j),j} + \sum_{k \in J_1, k \neq \ell} P_{(\ell,j),k} + q_{\ell j} \geq 1.$$

Putting it together with (C.6), we get for any $j \in J_1$ and for any $\ell \in J_2$, we have

$$x_j + q_{\ell j} \geq 1.$$

Therefore,

$$\sum_{j \in J_2} x_j + \sum_{j \in J_2} q_{\ell j} \geq m/2.$$

Now consider (C.5) for $\boldsymbol{h} = \boldsymbol{0}$. We have for $\ell \in J_1$,

$$x_\ell - \sum_{j \in J_2} q_{\ell j} \geq 0.$$

Moreover since $x_\ell \leq 1$, we conclude that,

$$\sum_{j \in J_2} q_{\ell j} \leq 1.$$

Therefore,

$$\sum_{j \in J_2} x_j \geq m/2 - 1.$$

Finally,

$$z_{\mathsf{Aff}} \geq \sum_{j=1}^{m} c_j x_j = \sum_{j \in J_2} x_j \geq m/2 - 1.$$

Now we consider the following affine solution

$$\begin{cases} x_i = 1 & \forall i = 1, \ldots, m-1 \\ x_m = 0 \end{cases}$$

$$\forall \boldsymbol{h} \in \mathcal{U}, \forall i \in J_1, \begin{cases} y_{ij}(\boldsymbol{h}) = 0 & \forall j \in J_2 \setminus \{m\} \\ y_{im}(\boldsymbol{h}) = -h_i + 1. \end{cases}$$

The above affine solution is feasible for the adjustable problem. In fact, The capacity constraints are verified on $\boldsymbol{x}$. The non-negativity constraints are verified on $\boldsymbol{y}(\boldsymbol{h})$. Demand is covered at each node by the first stage solution $\boldsymbol{x}$ except a node m. Demand at node m is covered in the second stage because

$$\sum_{i \in J_1} y_{im}(\boldsymbol{h}) = \sum_{i \in J_1} (1 - h_i) = m/2 - \sum_{i \in J_1} h_i \geq h_m,$$

where the inequality follows from the definition of $\mathcal{U}$. The cost of this affine solution is $m/2 - 1$. Therefore,

$$z_{\mathsf{Aff}}(\mathcal{U}) = m/2 - 1.$$

C.5 Proof of Lemma 3.7.1

The adjustable problem corresponding to the instance (3.26) is given by

$$\begin{aligned} z_{\mathsf{AR}}(\mathcal{U}) = \min_{\boldsymbol{x} \geq \boldsymbol{0}} \max_{\boldsymbol{h} \in \mathcal{U}} \min_{\boldsymbol{y}(\boldsymbol{h}) \geq \boldsymbol{0}} & \ \boldsymbol{e}^T \boldsymbol{y}(\boldsymbol{h}) \\ & \sum_{i=1}^{m} h_i (\boldsymbol{e} - \boldsymbol{e}_i) \boldsymbol{e}^T \boldsymbol{x} + \boldsymbol{y}(\boldsymbol{h}) \ \geq \ \boldsymbol{h}, \quad \forall \boldsymbol{h} \in \mathcal{U}. \end{aligned}$$

We introduce the new variable $\alpha = \boldsymbol{e}^T \boldsymbol{x}$. The problem is equivalent to

$$\begin{aligned} z_{\mathsf{AR}}(\mathcal{U}) = \min_{\alpha \geq 0} \max_{\boldsymbol{h} \in \mathcal{U}} \min_{\boldsymbol{y}(\boldsymbol{h}) \geq \boldsymbol{0}} & \ \boldsymbol{e}^T \boldsymbol{y}(\boldsymbol{h}) \\ & \alpha \sum_{i=1}^{m} \left(\sum_{j=1, j \neq i}^{m} h_j \right) \boldsymbol{e}_i + \boldsymbol{y}(\boldsymbol{h}) \ \geq \ \boldsymbol{h}, \quad \forall \boldsymbol{h} \in \mathcal{U}. \end{aligned}$$

Note that the first stage cost corresponding to α is zero and therefore we could choose the variable α to be arbitrary large in the optimal solution.

First remark that $z_{\mathsf{AR}}(\mathcal{U}) \geq 1$. In fact, by taking $\boldsymbol{h} = \boldsymbol{e}_1$, the first feasibility constraint implies $y_1(\boldsymbol{e}_1) \geq 1$ and therefore the cost of an adjustable solution is at least $\boldsymbol{e}^T \boldsymbol{y}(\boldsymbol{e}_1) \geq 1$. Hence, $z_{\mathsf{AR}}(\mathcal{U}) \geq 1$. Now let us construct a feasible solution of the adjustable problem with a cost equal to 1. Consider an extreme point $\boldsymbol{h}$ of $\mathcal{U}$. Therefore, $h_i \in \{0, 1\}$ for all $i \in [m]$. If $\boldsymbol{h} = \boldsymbol{0}$, the solution $\boldsymbol{y}(\boldsymbol{0}) = \boldsymbol{0}$ is feasible and its corresponding cost is 0. Otherwise, if $\boldsymbol{h}$ has at least one non-zero component, let say $h_i = 1$. Then, we consider the following solution $\boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{e}_i$ which is feasible. In fact, the i-th constraint is verified because

$$\alpha\left(\sum_{j=1, j\neq i}^{m} h_j\right) + y_i(\boldsymbol{h}) \geq y_i(\boldsymbol{h}) = 1 = h_i.$$

The other feasibility constraints are also verified because the first term in these constraints is non-zero and we could take α as large as needed. In particular, by just taking $\alpha = 1$, we have for any $k \neq i$,

$$\alpha\left(\sum_{j=1, j\neq k}^{m} h_j\right) + y_k(\boldsymbol{h}) \geq \alpha\left(\sum_{j=1, j\neq k}^{m} h_j\right) \geq \alpha h_i = 1 \geq h_k.$$

The corresponding cost is $\boldsymbol{e}^T \boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{e}^T \boldsymbol{e}_i = 1$. Hence, $z_{\mathsf{AR}}(\mathcal{U}) = 1$. Now let us compute the optimal affine solution. Consider an affine solution

$$\boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}.$$

We have for any $\boldsymbol{h} \in \mathcal{U}$,

$$z_{\mathsf{Aff}}(\mathcal{U}) \geq \boldsymbol{e}^T(\boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}).$$

Consider $\boldsymbol{h} = \boldsymbol{e}$. Then the above inequality implies,

$$z_{\mathsf{Aff}}(\mathcal{U}) \geq \sum_{i=1}^{m}\sum_{j=1}^{m} P_{ij} + \sum_{i=1}^{m} q_i. \tag{C.7}$$

Moreover, we know that $\boldsymbol{y}(\boldsymbol{h}) \geq \boldsymbol{0}$ for any $\boldsymbol{h} \in \mathcal{U}$. In particular, for all $i \in [m]$,

$$y_i(\boldsymbol{e} - \boldsymbol{e}_i) \geq 0.$$

Hence, for all $i \in [m]$,

$$\sum_{j=1, j\neq i}^{m} P_{ij} + q_i \geq 0. \tag{C.8}$$

Combining (C.7) and (C.8), we get

$$z_{\mathsf{Aff}}(\mathcal{U}) \geq \sum_{i=1}^{m} P_{ii}. \tag{C.9}$$

On the other hand, we have for $\boldsymbol{h} = \boldsymbol{0}$,

$$z_{\mathsf{Aff}}(\mathcal{U}) \geq \boldsymbol{e}^T \boldsymbol{y}(\boldsymbol{0}) = \sum_{i=1}^{m} q_i. \tag{C.10}$$

We know by feasibility of $\boldsymbol{y}(\boldsymbol{h})$ that for all $i \in [m]$,

$$\alpha \left(\sum_{j=1, j\neq i}^{m} h_j \right) + y_i(\boldsymbol{h}) \geq h_i.$$

In particular, for all $i \in [m]$, by taking $\boldsymbol{h} = \boldsymbol{e}_i$, the i-th constraint gives

$$y_i(\boldsymbol{e_i}) = P_{ii} + q_i \geq 1. \tag{C.11}$$

Therefore, from (C.9), (C.10) and (C.11), we conclude

$$z_{\mathsf{Aff}}(\mathcal{U}) \geq \sum_{i=1}^{m} (P_{ii} + q_i)/2 \geq \frac{m}{2}.$$

Finally, consider the following affine solution for any $\boldsymbol{h} \in \mathcal{U}$,

$$y_i(\boldsymbol{h}) = \begin{cases} h_i & \forall i = 1, \ldots, \frac{m}{2} \\ 1 - h_{i-\frac{m}{2}} & \forall i = \frac{m}{2} + 1, \ldots, m \end{cases}$$

where we assume m is even for the sake of simplicity. The above solution is feasible. In fact, for $i = 1, \ldots, \frac{m}{2}$, we have

$$\alpha \left(\sum_{j=1, j\neq i}^{m} h_j \right) + y_i(h) \geq y_i(h) = h_i$$

For $i = \frac{m}{2} + 1, \ldots, m$, if $h_{i-\frac{m}{2}} = 0$, then $y_i(\boldsymbol{h}) = 1$ and therefore the i-th constraint is verified.

If $h_{i-\frac{m}{2}} \neq 0$, therefore $\left(\sum_{j=1,j\neq i}^{m} h_j\right) \neq 0$ and by taking α sufficiently large, the i-th constraint is verified as well. Finally the cost of the proposed affine solution is

$$\boldsymbol{e}^T \boldsymbol{y}(\boldsymbol{h}) = m/2.$$

We conclude that

$$z_{\mathsf{Aff}}(\boldsymbol{\mathcal{U}}) = m/2.$$

Appendix D: Piecewise static policies

D.1 Proof of Lemma 6.3.5

First, note that for $h > 0$,

$$T(\mathcal{U},\boldsymbol{h}) = \left\{ \left(\frac{y_1}{h_1}, \frac{y_2}{h_2}, \ldots, \frac{y_m}{h_m}\right) \;\middle|\; (y_1, y_2, \ldots, y_m) \in T(\mathcal{U},\boldsymbol{e}) \right\}.$$

Then we can easily prove that $\kappa(T(\mathcal{U},\boldsymbol{h})) = \kappa(T(\mathcal{U},\boldsymbol{e}))$. In fact, let $\boldsymbol{x} \in \mathsf{conv}(T(\mathcal{U},\boldsymbol{h}))$. Then, $\sum_{i=1}^m x_i h_i \boldsymbol{e}_i \in \mathsf{conv}(T(\mathcal{U},\boldsymbol{e}))$. Therefore,

$$\frac{1}{\kappa(T(\mathcal{U},\boldsymbol{e}))} \cdot \left(\sum_{i=1}^m x_i h_i \boldsymbol{e}_i\right) \in T(\mathcal{U},\boldsymbol{e}).$$

Then,

$$\frac{1}{\kappa T((\mathcal{U},\boldsymbol{e}))} \cdot \boldsymbol{x} \in T(\mathcal{U},\boldsymbol{h}),$$

which implies,

$$\mathsf{conv}(T(\mathcal{U},\boldsymbol{h})) \subseteq \kappa(T(\mathcal{U},\boldsymbol{e})) \cdot T(\mathcal{U},\boldsymbol{h}),$$

and finally $\kappa(T(\mathcal{U},\boldsymbol{h})) \leq \kappa(T(\mathcal{U},\boldsymbol{e}))$. Similarly, we also have $\kappa(T(\mathcal{U},\boldsymbol{h})) \geq \kappa(T(\mathcal{U},\boldsymbol{e}))$. Now, it's sufficient to show that $\kappa(T(\mathcal{U},\boldsymbol{e})) = \sum_{i=1}^m \tau_i$. Let first show that

$$\mathsf{conv}(T(\mathcal{U},\boldsymbol{e})) = \left\{ (x_1, x_2, \ldots, x_m) \in [0,1]^m \;\middle|\; \sum_{i=1}^m \frac{x_i}{\tau_i} \leq 1 \right\}. \tag{D.1}$$

Let $\boldsymbol{x} \in \mathsf{conv}(T(\mathcal{U},\boldsymbol{e}))$. From Lemma 6.2.4, we have $\boldsymbol{x} = \sum_{i=1}^m \lambda_i a_i \boldsymbol{e}_i$, where $\sum_{i=1}^m \lambda_i = 1$, $\lambda_i \in [0,1]$ and $0 \leq a_i \leq \tau_i$, $\forall i \in [m]$. We have,

$$\sum_{i=1}^{m} \frac{x_i}{\tau_i} = \sum_{i=1}^{m} \lambda_i \cdot \frac{a_i}{\tau_i} \leq \sum_{i=1}^{m} \lambda_i = 1.$$

Conversely, let $\boldsymbol{x} \in \mathbb{R}_+^m$ such that,

$$\sum_{i=1}^{m} \frac{x_i}{\tau_i} \leq 1.$$

We have

$$\boldsymbol{x} = \sum_{j=1}^{m} \lambda_j a_j \boldsymbol{e}_j,$$

where for all $j \in [m]$,

$$\lambda_j = \frac{\frac{x_j}{\tau_j}}{\sum_{i=1}^{m} \frac{x_i}{\tau_i}} \qquad \text{and} \qquad a_j = \tau_j \sum_{i=1}^{m} \frac{x_i}{\tau_i}.$$

We have $\sum_{j=1}^{m} \lambda_j = 1$ and $a_j \leq \tau_j \ \forall j \in [m]$. Then, $\boldsymbol{x} \in \mathsf{conv}(T(\mathcal{U}, \boldsymbol{e}))$.

Now, we would like to find a lower bound for $\kappa(T(\mathcal{U}, \boldsymbol{e}))$. Let $\alpha \geq 1$ such that $\mathsf{conv}(T(\mathcal{U}, \boldsymbol{e})) \subseteq \alpha \cdot T(\mathcal{U}, \boldsymbol{e})$. From (D.1), we have

$$\left(\frac{\tau_1^2}{\sum_{i=1}^{m} \tau_i}, \frac{\tau_2^2}{\sum_{i=1}^{m} \tau_i}, \ldots, \frac{\tau_m^2}{\sum_{i=1}^{m} \tau_i} \right) \in \mathsf{conv}(T(\mathcal{U}, \boldsymbol{e}))$$

Then, there exists $\mathsf{diag}(\boldsymbol{x}) \in \mathcal{U}$ and $\boldsymbol{\mu} \in \mathbb{R}_+^m$, $\sum_{i=1}^{m} \mu_i = 1$, such that

$$\left(\frac{\tau_1^2}{\sum_{i=1}^{m} \tau_i}, \frac{\tau_2^2}{\sum_{i=1}^{m} \tau_i}, \ldots, \frac{\tau_m^2}{\sum_{i=1}^{m} \tau_i} \right) = \alpha \cdot \mathsf{diag}(\boldsymbol{x})^T \boldsymbol{\mu},$$

i.e. $\forall 1 \leq i \leq m$,

$$\frac{\tau_i^2}{\sum_{j=1}^{m} \tau_j} = \alpha \mu_i x_i$$

From Cauchy-Shwartz inequality we have,

$$\sum_{i=1}^{m} \frac{\tau_i^2}{\mu_i} \geq \left(\sum_{i=1}^{m} \tau_i \right)^2,$$

Then,

$$\alpha\left(\sum_{i=1}^{m}\tau_i\right)\left(\sum_{i=1}^{m}x_i\right)\geq\left(\sum_{i=1}^{m}\tau_i\right)^2,$$

i.e.

$$\alpha\left(\sum_{i=1}^{m}x_i\right)\geq\left(\sum_{i=1}^{m}\tau_i\right),$$

therefore,

$$\alpha\geq\sum_{i=1}^{m}\tau_i,$$

where the last inequality follows from $\sum_{i=1}^{m}x_i\leq 1$. To finish our proof we show that,

$$\mathsf{conv}(T(\mathcal{U},\boldsymbol{e}))\subseteq\left(\sum_{i=1}^{m}\tau_i\right)\cdot T(\mathcal{U},\boldsymbol{e}).$$

Let $\boldsymbol{x}\in\mathsf{conv}(T(\mathcal{U},\boldsymbol{e}))$, we have from (D.1),

$$\sum_{i=1}^{m}\frac{x_i}{\tau_i}\leq 1.$$

For all $1\leq j\leq m$, let define,

$$\mu_j=\frac{\frac{x_j}{\tau_j}}{\sum_{i=1}^{m}\frac{x_i}{\tau_i}}\qquad\text{and}\qquad b_j=\tau_j\frac{\sum_{i=1}^{m}\frac{x_i}{\tau_i}}{\sum_{i=1}^{m}\tau_i}.$$

Then

$$\boldsymbol{x}=\left(\sum_{i=1}^{m}\tau_i\right)\cdot\mathsf{diag}(\boldsymbol{b})^T\boldsymbol{\mu}$$

We have $\forall j\in[m]$,

$$b_j\leq\frac{\tau_j}{\sum_{i=1}^{m}\tau_i}\leq\tau_j$$

where the second inequality holds because $\sum_{i=1}^{m}\tau_i\geq 1$. Furthermore,

$$\sum_{j=1}^{m}b_j=\sum_{i=1}^{m}\frac{x_i}{\tau_i}\leq 1.$$

Therefore, $\mathsf{diag}(\boldsymbol{b}) \in \mathcal{U}$. Since $\sum_{j=1}^{n} \mu_j = 1$, $\mathsf{diag}(\boldsymbol{b})^T \boldsymbol{\mu} \in T(\mathcal{U}, \boldsymbol{e})$. We conclude that

$$\boldsymbol{x} \in \left(\sum_{i=1}^{m} \tau_i \right) \cdot T(\mathcal{U}, \boldsymbol{e}).$$

Appendix E: Piecewise affine policies

E.1 Proof of Theorem 4.2.3

Proof. Let $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}(\hat{\boldsymbol{h}}), \hat{\boldsymbol{h}} \in \hat{\mathcal{U}})$ be an optimal solution for $z_{\mathsf{AR}}(\hat{\mathcal{U}})$. For each $\boldsymbol{h} \in \mathcal{U}$, let $\tilde{\boldsymbol{y}}(\boldsymbol{h}) = \hat{\boldsymbol{y}}(\hat{\boldsymbol{h}})$ where $\hat{\boldsymbol{h}} \in \hat{\mathcal{U}}$ dominates $\boldsymbol{h}$. Therefore, for any $\boldsymbol{h} \in \mathcal{U}$,

$$\boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}\tilde{\boldsymbol{y}}(\boldsymbol{h}) = \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}\hat{\boldsymbol{y}}(\hat{\boldsymbol{h}}) \geq \hat{\boldsymbol{h}} \geq \boldsymbol{h},$$

i.e., $(\hat{\boldsymbol{x}}, \tilde{\boldsymbol{y}}(\boldsymbol{h}), \boldsymbol{h} \in \mathcal{U})$ is a feasible solution for $z_{\mathsf{AR}}(\mathcal{U})$. Therefore,

$$z_{\mathsf{AR}}(\mathcal{U}) \leq \boldsymbol{c}^T\hat{\boldsymbol{x}} + \max_{\boldsymbol{h}\in\mathcal{U}} \boldsymbol{d}^T\tilde{\boldsymbol{y}}(\boldsymbol{h}) \leq \boldsymbol{c}^T\hat{\boldsymbol{x}} + \max_{\hat{\boldsymbol{h}}\in\hat{\mathcal{U}}} \boldsymbol{d}^T\hat{\boldsymbol{y}}(\hat{\boldsymbol{h}}) = z_{\mathsf{AR}}(\hat{\mathcal{U}}).$$

Conversely, let $(\boldsymbol{x}^*, \boldsymbol{y}^*(\boldsymbol{h}), \boldsymbol{h} \in \mathcal{U})$ be an optimal solution of $z_{\mathsf{AR}}(\mathcal{U})$. Then, for any $\hat{\boldsymbol{h}} \in \hat{\mathcal{U}}$, since $\frac{\hat{\boldsymbol{h}}}{\beta} \in \mathcal{U}$, we have,

$$\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{y}^*\left(\frac{\hat{\boldsymbol{h}}}{\beta}\right) \geq \frac{\hat{\boldsymbol{h}}}{\beta},$$

Therefore, $\left(\beta\boldsymbol{x}^*, \beta\boldsymbol{y}^*\left(\frac{\hat{\boldsymbol{h}}}{\beta}\right), \hat{\boldsymbol{h}} \in \mathcal{U}\right)$ is feasible for $\Pi_{\mathsf{AR}}(\hat{\mathcal{U}})$. Therefore,

$$z_{\mathsf{AR}}(\hat{\mathcal{U}}) \leq \boldsymbol{c}^T\beta\boldsymbol{x}^* + \max_{\hat{\boldsymbol{h}}\in\hat{\mathcal{U}}} \boldsymbol{d}^T\beta\boldsymbol{y}^*\left(\frac{\hat{\boldsymbol{h}}}{\beta}\right) \leq \beta \cdot \left(\boldsymbol{c}^T\boldsymbol{x}^* + \max_{\boldsymbol{h}\in\mathcal{U}} \boldsymbol{d}^T\boldsymbol{y}^*(\boldsymbol{h})\right) = \beta \cdot z_{\mathsf{AR}}(\mathcal{U}).$$

□

E.2 Proof of Lemma 4.2.4

Proof. a) Suppose there exists β and $\boldsymbol{v} \in \mathcal{U}$ such that $\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\,(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{v})$ dominates $\mathcal{U}$. Consider $\boldsymbol{h} \in \mathcal{U}$. Since $\hat{\mathcal{U}}$ dominates $\mathcal{U}$, there exists $\alpha_1, \alpha_2, \ldots, \alpha_{m+1} \geq 0$ with $\alpha_1 + \ldots + \alpha_{m+1} = 1$

such that

$$h_i \le \beta\left(\alpha_i + \alpha_{m+1} v_i\right),\ \forall i = 1, \ldots, m. \tag{E.1}$$

Let

$$I(\boldsymbol{h}) = \left\{ i \in [m] \;\middle|\; h_i - \beta v_i \ge 0 \right\}.$$

Then,

$$\begin{aligned}
\sum_{i=1}^{m} \left(h_i - \beta v_i\right)^+ &= \sum_{i \in I(\boldsymbol{h})} h_i - \beta \sum_{i \in I(\boldsymbol{h})} v_i \\
&\le \sum_{i \in I(\boldsymbol{h})} \beta\left(\alpha_i + \alpha_{m+1} v_i\right) - \beta \sum_{i \in I(\boldsymbol{h})} v_i \\
&= \beta \sum_{i \in I(\boldsymbol{h})} \alpha_i + \left(\alpha_{m+1} - 1\right) \beta \sum_{i \in I(\boldsymbol{h})} v_i \\
&\le \beta,
\end{aligned}$$

where the first inequality follows from (E.1) and the last inequality holds because $\alpha_{m+1} - 1 \le 0$, $v_i \ge 0$, $\beta \ge 0$ and $\sum_{i \in I(\boldsymbol{h})} \alpha_i \le 1$. We conclude that

$$\frac{1}{\beta} \sum_{i=1}^{m} \left(h_i - \beta v_i\right)^+ \le 1.$$

b) Now, suppose there exists β and $\boldsymbol{v} \in \mathcal{U}$ such that $\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\left(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{v}\right)$ dominates $\mathcal{U}$. For any $\boldsymbol{h} \in \mathcal{U}$, let

$$\hat{\boldsymbol{h}} = \sum_{i=1}^{m} \left(h_i - \beta v_i\right)^+ \boldsymbol{e}_i + \beta \boldsymbol{v}.$$

Then for all $i = 1, \ldots, m$,

$$\begin{aligned}
\hat{h}_i &= \left(h_i - \beta v_i\right)^+ + \beta v_i \\
&\ge \left(h_i - \beta v_i\right) + \beta v_i \ge h_i.
\end{aligned}$$

Therefore, $\hat{\boldsymbol{h}}$ dominates $\boldsymbol{h}$. Moreover,

$$\hat{\boldsymbol{h}} = 2\beta\left(\sum_{i=1}^{m}\frac{(h_i-\beta v_i)^+}{2\beta}\boldsymbol{e}_i + \frac{1}{2}\boldsymbol{v}\right) \in 2\beta\cdot\mathsf{conv}\left(\boldsymbol{0},\boldsymbol{e}_1,\ldots,\boldsymbol{e}_m,\boldsymbol{v}\right),$$

because

$$\frac{1}{\beta}\sum_{i=1}^{m}(h_i-\beta v_i)^+ \leq 1.$$

Therefore, $2\beta\cdot\mathsf{conv}\left(\boldsymbol{0},\boldsymbol{e}_1,\ldots,\boldsymbol{e}_m,\boldsymbol{v}\right)$ dominates $\mathcal{U}$ and consequently

$2\beta\cdot\mathsf{conv}\left(\boldsymbol{e}_1,\ldots,\boldsymbol{e}_m,\boldsymbol{v}\right)$ dominates $\mathcal{U}$ as well. □

E.3 Proof of Lemma 4.3.2

Proof. Suppose $k\in[m]$. Let us consider

$$\boldsymbol{h}\in\operatorname*{argmax}_{\boldsymbol{h}\in\mathcal{U}}\sum_{i=1}^{k}h_i.$$

Without loss of generality, we can suppose that $h_i=0$ for $i=k+1,\ldots,m$. Denote, $\mathcal{S}_k$ the set of permutations of $\{1,2,\ldots,k\}$. We define $\boldsymbol{h}^\sigma\in\mathbb{R}_+^m$ such that $h_i^\sigma=h_{\sigma(i)}$ for $i=1,\ldots,k$ and $h_i^\sigma=0$ otherwise. Since $\mathcal{U}$ is a permutation invariant set, we have $\boldsymbol{h}^\sigma\in\mathcal{U}$ for any $\sigma\in\mathcal{S}_k$. The convexity of $\mathcal{U}$ implies that

$$\frac{1}{k!}\sum_{\sigma\in\mathcal{S}_k}\boldsymbol{h}^\sigma\in\mathcal{U}.$$

We have,

$$\sum_{\sigma\in\mathcal{S}_k}h_i^\sigma=\begin{cases}(k-1)!\cdot\sum_{j=1}^{k}h_j & \text{if } i=1,\ldots,k\\ 0 & \text{otherwise,}\end{cases}$$

and $\sum_{j=1}^{k}h_j=k\cdot\gamma(k)$ by definition. Therefore,

$$\frac{1}{k!}\sum_{\sigma\in\mathcal{S}_k}\boldsymbol{h}^\sigma=\gamma(k)\cdot\sum_{i=1}^{k}\boldsymbol{e}_i\in\mathcal{U}.$$

□

E.4 Proof of Lemma 4.3.3

Proof. Consider, $\tilde{\boldsymbol{h}} \in \mathcal{U}$ an optimal solution for the maximization problem in (4.9) for fixed β. We will construct $\boldsymbol{h}^* \in \mathcal{U}$ another optimal solution of (4.9) that verifies the properties in the lemma. First, denote $I = \{i \mid \tilde{h}_i > \beta\gamma\}$ and $|I| = k$. Since, $\mathcal{U}$ is permutation invariant, we can suppose without loss of generality that $I = \{1, 2, \ldots, k\}$. We define,

$$h_i^* = \begin{cases} \gamma(k) & \text{if } i = 1, \ldots, k \\ 0 & \text{otherwise.} \end{cases}$$

From Lemma 4.3.2, we have $\boldsymbol{h}^* \in \mathcal{U}$. Moreover,

$$\begin{aligned} \sum_{i=1}^{m} (\tilde{h}_i - \beta\gamma)^+ = \sum_{i=1}^{k} \tilde{h}_i - \beta\gamma k \leq k \cdot \gamma(k) - \beta\gamma k \\ = \sum_{i=1}^{k} (\gamma(k) - \beta\gamma) = \sum_{i=1}^{k} (h_i^* - \beta\gamma) \\ \leq \sum_{i=1}^{k} (h_i^* - \beta\gamma)^+ = \sum_{i=1}^{m} (h_i^* - \beta\gamma)^+ \end{aligned}$$

where the first inequality follows from the definition of the coefficients $\gamma(.)$. Therefore, $\boldsymbol{h}^*$ and $\tilde{\boldsymbol{h}}$ have the same objective value in (4.9) and consequently $\boldsymbol{h}^*$ is also optimal for the maximization problem (4.9). Moreover, from the first inequality, we have $\gamma(k) - \beta\gamma > 0$, i.e., $|\{i \mid h_i^* > \beta\gamma\}| = k$. Therefore, $\boldsymbol{h}^*$ verifies the properties of the lemma. □

E.5 Proof of Proposition 4.3.9

Proof. To prove that $\hat{\mathcal{U}}$ dominates $\mathcal{U}$, it is sufficient to take $\boldsymbol{h}$ in the boundaries of $\mathcal{U}$, i.e.,

$$a \sum_{i=1}^{m} h_i \sum_{j=1}^{m} h_j + (1 - a) \sum_{i=1}^{m} h_i^2 = 1, \tag{E.2}$$

and find $\alpha_1, \alpha_2, \ldots, \alpha_{m+1}$ nonnegative reals with $\sum_{i=1}^{m+1} \alpha_i = 1$ such that for all $i \in [m]$,

$$h_i \leq \beta \left(\alpha_i + \gamma \alpha_{m+1}\right).$$

By taking all h_i equal in (E.2), we get

$$\gamma = \frac{1}{\sqrt{(am^2 + (1-a)m)}}.$$

We choose for $i \in [m]$,

$$\alpha_i = \frac{1}{2}\left((1-a)h_i^2 + a h_i \sum_{j=1}^{m} h_j\right)$$

and $\alpha_{m+1} = \frac{1}{2}$. First, we have $\sum_{i=1}^{m+1} \alpha_i = 1$ and for all $i \in [m]$,

$$\begin{aligned}
\beta\left(\alpha_i + \gamma \alpha_{m+1}\right) &= \frac{\beta}{2}\left((1-a)h_i^2 + a h_i \sum_{j=1}^{m} h_j + \frac{1}{\sqrt{am^2 + (1-a)m}}\right) \\
&\geq \frac{\beta}{2}\left((1-a)h_i^2 + \frac{1}{\sqrt{am^2 + (1-a)m}} + a h_i\right) \\
&\geq \frac{\beta}{2}\left(2\left(\frac{(1-a)}{\sqrt{am^2 + (1-a)m}}\right)^{\frac{1}{2}} h_i + a h_i\right) = h_i
\end{aligned}$$

where the first inequality holds because $\sum_{j=1}^{m} h_j \geq 1$ which is a direct consequence of $\boldsymbol{h}^T \Sigma \boldsymbol{h} = 1$ and $a \leq 1$. The second one follows from the inequality of arithmetic and geometric means (AM-GM inequality). Finally, we can verify by case analysis on the values of a that

$$\left(\frac{a}{2} + \frac{(1-a)^{\frac{1}{2}}}{(am^2 + (1-a)m)^{\frac{1}{4}}}\right)^{-1} = O\left(m^{\frac{2}{5}}\right).$$

In fact, denote $H(m) = \left(\frac{a}{2} + \frac{(1-a)^{\frac{1}{2}}}{(am^2+(1-a)m)^{\frac{1}{4}}}\right)^{-1} = O\left(a + \frac{1}{(am^2+m)^{\frac{1}{4}}}\right)^{-1}$

Case1: $a = O(\frac{1}{m})$**.** We have $\left(am^2 + m\right)^{\frac{1}{4}} = O(m^{\frac{1}{4}})$. Then $H(m) = O(m^{\frac{1}{4}}) = O(m^{\frac{2}{5}})$.

Case2: $a = \Omega(m^{\frac{-2}{5}})$**.** We have $H(m) = O(a^{-1}) = O(m^{\frac{2}{5}})$.

Case3: $a = O(m^{\frac{-2}{5}})$ **and** $a = \Omega(\frac{1}{m})$**.** We have $\left(am^2 + m\right)^{\frac{1}{4}} = O(m^{\frac{2}{5}})$. Then,

$$a + \frac{1}{(am^2 + m)^{\frac{1}{4}}} = \Omega(\frac{1}{m}) + \Omega(m^{\frac{-2}{5}}) = \Omega(m^{\frac{-2}{5}}).$$

Therefore, $H(m) = O(m^{\frac{2}{5}})$. □

E.6 Proof of Proposition 4.3.10

Proof. To prove that $\hat{\mathcal{U}}$ dominates $\mathcal{U}$, it is sufficient to take $\boldsymbol{h}$ in the boundaries of $\mathcal{U}$, i.e., $\sum_{i=1}^{m} h_i = k$ and find $\alpha_1, \alpha_2, \ldots, \alpha_{m+1}$ non-negative reals with $\sum_{i=1}^{m+1} \alpha_i = 1$ such that for all $i \in [m]$,

$$h_i \leq \beta \left(\alpha_i + \frac{k}{m}\alpha_{m+1}\right).$$

<u>First case:</u> If $\beta = k$, we choose $\alpha_i = \frac{h_i}{k}$ for $i \in [m]$ and $\alpha_{m+1} = 0$. We have $\sum_{i=1}^{m+1} \alpha_i = 1$ and for all $i \in [m]$,

$$\beta \left(\alpha_i + \frac{k}{m}\alpha_{m+1}\right) = k\frac{h_i}{k} \geq h_i.$$

<u>Second case:</u> If $\beta = \frac{m}{k}$, we choose $\alpha_i = 0$ for $i \in [m]$ and $\alpha_{m+1} = 1$. We have $\sum_{i=1}^{m+1} \alpha_i = 1$ and for all $i \in [m]$,

$$\beta \left(\alpha_i + \frac{k}{m}\alpha_{m+1}\right) = 1 \geq h_i.$$

□

E.7 Proof of Lemma 4.3.11

Proof. Consider the following simplex

$$\hat{\mathcal{U}} = \mathsf{conv}\left(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \frac{1}{\sqrt{m}}\boldsymbol{e}\right)$$

It is clear that $\hat{\mathcal{U}}$ dominates $\mathcal{U}$ since $\frac{1}{\sqrt{m}}\boldsymbol{e}$ dominates all the extreme points $\boldsymbol{v}_j$ for $j \in [N]$. Moreover, by the convexity of $\mathcal{U}$, we have $\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{v}_j = \frac{\binom{m-1}{r-1}}{\sqrt{m}\binom{m}{r}}\boldsymbol{e} = \frac{r}{m\sqrt{m}}\boldsymbol{e} \in \mathcal{U}$. Denote $\beta = \frac{m}{r}$. Hence, for all $i \in [m]$

$$\boldsymbol{e}_i = \beta \underbrace{\left(\frac{1}{\beta}\cdot \boldsymbol{e}_i + (1-\frac{1}{\beta})\cdot \boldsymbol{0}\right)}_{\in\mathcal{U}} \qquad \text{and} \qquad \frac{1}{\sqrt{m}}\boldsymbol{e} = \beta \cdot \underbrace{\frac{r}{m\sqrt{m}}}_{\in\mathcal{U}}\boldsymbol{e}\,.$$

Therefore, $\hat{\mathcal{U}} \subseteq \beta\cdot\mathcal{U}$ and from Theorem 4.2.3, we conclude that our policy gives a β-approximation to the adjustable problem (1.1) where $\beta = \frac{m}{\lceil m-\sqrt{m}\rceil} = O(1+\frac{1}{\sqrt{m}})$. □

E.8 Proof of Lemma 4.3.12

Proof. First, let us prove that $z_{\mathsf{AR}}(\mathcal{U}) \leq 1$. It is sufficient to define an adjustable solution only for the extreme points of $\mathcal{U}$ because the constraints are linear. We define the following solution for all $i = 1,\ldots,m$ and for all $j = 1,\ldots,N$

$$\boldsymbol{x} = \boldsymbol{0}, \qquad \boldsymbol{y}(\boldsymbol{0}) = \boldsymbol{0}, \qquad \boldsymbol{y}(\boldsymbol{e}_i) = \boldsymbol{e}_i, \qquad \boldsymbol{y}(\boldsymbol{v}_j) = \frac{1}{m}\boldsymbol{e}.$$

We have $\boldsymbol{B}\boldsymbol{y}(\boldsymbol{0}) = \boldsymbol{0}$. For $i \in [m]$

$$\boldsymbol{B}\boldsymbol{y}(\boldsymbol{e}_i) = \boldsymbol{e}_i + \frac{1}{\sqrt{m}}(\boldsymbol{e}-\boldsymbol{e}_i) \geq \boldsymbol{e}_i$$

and for $j \in [N]$

$$\boldsymbol{B}\boldsymbol{y}(\boldsymbol{v}_j) = \frac{1}{m}\boldsymbol{B}\boldsymbol{e} = \left(\frac{1}{m} + \frac{m-1}{m\sqrt{m}}\right)\boldsymbol{e} \geq \frac{1}{\sqrt{m}}\boldsymbol{e} \geq \boldsymbol{v}_j.$$

Therefore, the solution defined above is feasible. Moreover, the cost of our feasible solution is 1 because for all $i \in [m]$ and $j \in [N]$, we have

$$\boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{e}_i) = \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{v}_j) = 1.$$

Hence, $z_{\mathsf{AR}}(\mathcal{U}) \leq 1$. Now, it is sufficient to prove that $z_{\mathsf{Aff}}(\mathcal{U}) = \Omega(\sqrt{m})$. First, $\tilde{\boldsymbol{x}} = \frac{1}{\sqrt{m}}\boldsymbol{e}$ and $\boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{0}$ for any $\boldsymbol{h} \in \mathcal{U}$ is a feasible static solution (which is a special case of an affine solution). In fact,

$$\boldsymbol{A}\tilde{\boldsymbol{x}} = \frac{1}{\sqrt{m}}\boldsymbol{A}\boldsymbol{e} = \left(\frac{1}{\sqrt{m}} + \frac{m-1}{m}\right)\boldsymbol{e} \geq \boldsymbol{e} \geq \boldsymbol{h} \qquad \forall \boldsymbol{h} \in \mathcal{U}$$

where the last inequality holds because $\mathcal{U} \subseteq [0,1]^m$. Moreover, the cost of this static solution is

$$\boldsymbol{c}^T\tilde{\boldsymbol{x}} = \frac{\sqrt{m}}{15}.$$

Hence,

$$z_{\mathsf{Aff}}(\mathcal{U}) \leq \frac{\sqrt{m}}{15}. \tag{E.3}$$

Our instance is "a permuted instance", i.e. $\mathcal{U}$ is permutation invariant, $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric and $\boldsymbol{c}$ and $\boldsymbol{d}$ are proportional to $\boldsymbol{e}$. Hence, from Lemma 8 and Lemma 7 in Bertsimas and Goyal [18], for any optimal solution $\boldsymbol{x}^*_{\mathsf{Aff}}, \boldsymbol{y}^*_{\mathsf{Aff}}(\boldsymbol{h})$ of the affine problem, we can construct another optimal affine solution that is "symmetric" and have the same stage cost. In particular, there exists an optimal solution for the affine problem of the following form $\boldsymbol{x} = \alpha\boldsymbol{e}$, $\boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}$ for $\boldsymbol{h} \in \mathcal{U}$ where

$$\boldsymbol{P} = \begin{pmatrix} \theta & \mu & \dots & \mu \\ \mu & \theta & \dots & \mu \\ \vdots & \vdots & \ddots & \vdots \\ \mu & \mu & \dots & \theta \end{pmatrix} \tag{E.4}$$

$\boldsymbol{q} = \lambda\boldsymbol{e}$, $\boldsymbol{c}^T\boldsymbol{x} = \boldsymbol{c}^T\boldsymbol{x}^*_{\mathsf{Aff}}$ and $\max_{\boldsymbol{h}\in\mathcal{U}} \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{h}) = \max_{\boldsymbol{h}\in\mathcal{U}} \boldsymbol{d}^T\boldsymbol{y}^*_{\mathsf{Aff}}(\boldsymbol{h})$. We have $\boldsymbol{x} \geq \boldsymbol{0}$ and $\boldsymbol{y}(\boldsymbol{0}) = \lambda\boldsymbol{e} \geq \boldsymbol{0}$ hence

$$\lambda \geq 0 \qquad \text{and} \qquad \alpha \geq 0. \tag{E.5}$$

<u>**Claim:**</u> $\alpha \geq \frac{1}{24\sqrt{m}}$ For a sake of contradiction, suppose that $\alpha > \frac{1}{24\sqrt{m}}$. We know that

$$z_{\mathsf{Aff}}(\mathcal{U}) \geq \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{0}) = \frac{\alpha}{15}m + \lambda m. \tag{E.6}$$

Case 1: If $\lambda \geq \frac{1}{12\sqrt{m}}$, then from (E.6) and $\alpha \geq 0$, we have $z_{\mathsf{Aff}}(\mathcal{U}) \geq \frac{\sqrt{m}}{12}$. Contradiction with (E.3).

Case 2: If $\lambda \leq \frac{1}{12\sqrt{m}}$. We have

$$\boldsymbol{y}(\boldsymbol{e}_1) = (\theta + \lambda)\boldsymbol{e}_1 + (\mu + \lambda)(\boldsymbol{e} - \boldsymbol{e}_1).$$

By feasibility of the solution, we have $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{e}_1) \geq \boldsymbol{e}_1$, hence

$$\theta + \lambda + \alpha \left(\frac{m-1}{\sqrt{m}} + 1 \right) + \frac{1}{\sqrt{m}}(m-1)(\mu + \lambda) \geq 1$$

Therefore $\theta + \lambda + \alpha \left(\frac{m-1}{\sqrt{m}} + 1 \right) \geq \frac{1}{2}$ or $\frac{1}{\sqrt{m}}(m-1)(\mu + \lambda) \geq \frac{1}{2}$.

Case 2.1: Suppose $\frac{1}{\sqrt{m}}(m-1)(\mu + \lambda) \geq \frac{1}{2}$. Therefore,

$$z_{\mathsf{Aff}}(\mathcal{U}) \geq \boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{e}_1) = \theta + \lambda + (m-1)(\mu + \lambda) \geq \frac{\sqrt{m}}{2}. \quad \text{(Contradiction with (E.3))}$$

where the last inequality holds because $\theta + \lambda \geq 0$ as $\boldsymbol{y}(\boldsymbol{e}_1) \geq \boldsymbol{0}$.

Case 2.2: Now suppose we have the other inequality i.e. $\theta + \lambda + \alpha \left(\frac{m-1}{\sqrt{m}} + 1 \right) \geq \frac{1}{2}$. Recall that we have $\lambda \leq \frac{1}{12\sqrt{m}}$ and we know that $\alpha < \frac{1}{24\sqrt{m}}$. Therefore,

$$\theta \geq \frac{1}{2} - \frac{1}{12\sqrt{m}} - \frac{1}{24\sqrt{m}} \left(\frac{m-1}{\sqrt{m}} + 1 \right) = \frac{11}{24} - \frac{3}{24\sqrt{m}} + \frac{1}{24m} \geq \frac{11}{24} - \frac{3}{24} = \frac{1}{3}.$$

We have,

$$\boldsymbol{y}(\boldsymbol{v}_1) = \frac{1}{\sqrt{m}} \left((\theta + (r-1)\mu)(\boldsymbol{e}_1 + \ldots \boldsymbol{e}_r) + r\mu(\boldsymbol{e} - (\boldsymbol{e}_1 + \ldots \boldsymbol{e}_r)) \right) + \lambda \boldsymbol{e}.$$

In particular we have ,

$$z_{\mathsf{Aff}}(\mathcal{U}) \geq \boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{v}_1) = \frac{r}{\sqrt{m}}(\theta + (m-1)\mu) + \lambda m$$
$$\geq \frac{r}{\sqrt{m}}\left(\frac{1}{3} + (m-1)\mu\right). \tag{E.7}$$

where the last inequality follows from $\lambda \geq 0$ and $\theta \geq \frac{1}{3}$.

Case 2.2.1: If $\mu \geq 0$ then from (E.7)

$$z_{\mathsf{Aff}}(\mathcal{U}) \geq \frac{r}{3\sqrt{m}} \geq \frac{m - \sqrt{m}}{3\sqrt{m}} \geq \frac{\sqrt{m}}{6} \quad \text{for } m \geq 4 \quad \text{(Contradiction with (E.3))}$$

Case 2.2.2: Now suppose that $\mu < 0$, by non-negativity of $\boldsymbol{y}(\boldsymbol{v}_1)$ we have

$$\frac{r}{\sqrt{m}}\mu + \lambda \geq 0$$

i.e.

$$\mu \geq \frac{-\lambda\sqrt{m}}{r}$$

and from (E.7)

$$\begin{aligned} z_{\mathsf{Aff}}(\mathcal{U}) &\geq \frac{r}{\sqrt{m}}\left(\frac{1}{3} + (m-1)\mu\right) \\ &\geq \frac{r}{\sqrt{m}}\left(\frac{1}{3} - \lambda\sqrt{m}\frac{m-1}{r}\right) \\ &\geq \frac{r}{\sqrt{m}}\left(\frac{1}{3} - \frac{1}{12}\frac{m-1}{r}\right) \geq \frac{r}{\sqrt{m}}\left(\frac{1}{3} - \frac{1}{6}\right) \quad \text{for } m \geq 4. \\ &\geq \frac{\sqrt{m}}{12} \quad \text{(Contradiction with (E.3))} \end{aligned}$$

We conclude that $\alpha \geq \frac{1}{24\sqrt{m}}$ and consequently

$$z_{\mathsf{Aff}}(\mathcal{U}) \geq \boldsymbol{c}^T \boldsymbol{x} = \frac{\alpha m}{15} \geq \frac{\sqrt{m}}{360} = \Omega(\sqrt{m}).$$

Hence,

$$z_{\mathsf{Aff}}(\mathcal{U}) = \Omega(\sqrt{m}) \cdot z_{\mathsf{AR}}(\mathcal{U}).$$

$\boldsymbol{c}^T \boldsymbol{x} = \boldsymbol{c}^T \boldsymbol{x}^*_{\mathsf{Aff}}$ Moreover, for any optimal affine solution, the cost of the first-stage affine solution $\boldsymbol{x}^*_{\mathsf{Aff}}$ is $\Omega(\sqrt{m})$ away from the optimal adjustable problem (1.1), i.e. $\boldsymbol{c}^T \boldsymbol{x}^*_{\mathsf{Aff}} = \boldsymbol{c}^T \boldsymbol{x} = \Omega(\sqrt{m}) \cdot z_{\mathsf{AR}}(\mathcal{U})$. □

E.9 Proof of Theorem 4.5.1

Proof. Let us find the order of the left hand side ratio in inequality (4.25). We have,

$$\begin{aligned}
\frac{\binom{\sqrt{m}}{m^\epsilon} \cdot \binom{m-m^\epsilon}{\sqrt{m}-m^\epsilon}}{\binom{m}{\sqrt{m}}} &= \frac{(\sqrt{m})! \times (m-m^\epsilon)! \times (m-\sqrt{m})! \times (\sqrt{m})!}{(\sqrt{m}-m^\epsilon)! \times (m^\epsilon)! \times m! \times (\sqrt{m}-m^\epsilon)! \times (m-\sqrt{m})!} \\
&= \left(\frac{(\sqrt{m})!}{(\sqrt{m}-m^\epsilon)!}\right)^2 \cdot \frac{(m-m^\epsilon)!}{(m^\epsilon)! \times m!}.
\end{aligned}$$

By Stirling's approximation, we have

$$\left(\sqrt{m}\right)! = \Theta\left(m^{\frac{1}{4}}\left(\frac{\sqrt{m}}{e}\right)^{\sqrt{m}}\right).$$

$$\left(\sqrt{m}-m^{\epsilon}\right)! = \Theta\left((\sqrt{m}-m^{\epsilon})^{\frac{1}{2}}\left(\frac{\sqrt{m}-m^{\epsilon}}{e}\right)^{\sqrt{m}-m^{\epsilon}}\right).$$

$$(m-m^{\epsilon})! = \Theta\left((m-m^{\epsilon})^{\frac{1}{2}}\left(\frac{m-m^{\epsilon}}{e}\right)^{m-m^{\epsilon}}\right).$$

$$(m)! = \Theta\left(m^{\frac{1}{2}}\left(\frac{m}{e}\right)^{m}\right).$$

$$(m^{\epsilon})! = \Theta\left(m^{\frac{1}{2}\epsilon}\left(\frac{m^{\epsilon}}{e}\right)^{m^{\epsilon}}\right).$$

All together,

$$\frac{\binom{\sqrt{m}}{m^{\epsilon}}\cdot\binom{m-m^{\epsilon}}{\sqrt{m}-m^{\epsilon}}}{\binom{m}{\sqrt{m}}} = \Theta\left(\frac{\left(\sqrt{m}\right)^{2\sqrt{m}}\cdot(m-m^{\epsilon})^{(m-m^{\epsilon})}}{m^{\frac{1}{2}\epsilon}\cdot\left(\sqrt{m}-m^{\epsilon}\right)^{2(\sqrt{m}-m^{\epsilon})}\cdot m^{m}\cdot m^{\epsilon m^{\epsilon}}}\right).$$

We have

$$(m-m^{\epsilon})^{(m-m^{\epsilon})} = \Theta\left(m^{(m-m^{\epsilon})}\cdot e^{-m^{\epsilon}+\frac{m^{2\epsilon}}{m}}\right),$$

and

$$\left(\sqrt{m}-m^{\epsilon}\right)^{2(\sqrt{m}-m^{\epsilon})} = \Theta\left(\left(\sqrt{m}\right)^{2(\sqrt{m}-m^{\epsilon})}\cdot e^{-2m^{\epsilon}+2\frac{m^{2\epsilon}}{\sqrt{m}}}\right),$$

WLOG, we can suppose that $\epsilon < \frac{1}{4}$, therefore

$$\begin{aligned}\frac{\binom{\sqrt{m}}{m^{\epsilon}}\cdot\binom{m-m^{\epsilon}}{\sqrt{m}-m^{\epsilon}}}{\binom{m}{\sqrt{m}}} &= \Theta\left(\frac{e^{m^{\epsilon}-2\frac{m^{2\epsilon}}{\sqrt{m}}+\frac{m^{2\epsilon}}{m}}}{m^{\epsilon m^{\epsilon}+\frac{1}{2}\epsilon}}\right)\\ &= \Theta\left(\frac{e^{m^{\epsilon}}}{m^{\epsilon m^{\epsilon}+\frac{1}{2}\epsilon}}\right).\end{aligned}$$

We have,

$$\Theta\left(\frac{Q(m)e^{m^\epsilon}}{m^{\epsilon m^\epsilon+\frac{1}{2}\epsilon}}\right) \geq 1,$$

but the later inequality contradicts

$$\lim_{m\to\infty} \frac{Q(m)e^{m^\epsilon}}{m^{\epsilon m^\epsilon+\frac{1}{2}\epsilon}} = 0.$$

□

E.10 Domination for non-permutation invariant sets

Propostion E.10.1. Suppose Algorithm 3 returns β and $\boldsymbol{v}$ for some uncertainty set $\mathcal{U}$. Then the set (4.28) is a dominating set for $\mathcal{U}$.

Proof. Suppose Algorithm 3 returns β and $\boldsymbol{v}$,then the inequality (4.5) is verified, namely,

$$\frac{1}{\beta}\sum_{i=1}^{m}(h_i - \beta v_i)^+ \leq 1,\ \forall \boldsymbol{h} \in \mathcal{U}.$$

Recall the dominating point (4.4)

$$\hat{\boldsymbol{h}}(\boldsymbol{h}) = \beta\boldsymbol{v} + (\boldsymbol{h} - \beta\boldsymbol{v})_+.$$

We have

$$\hat{\boldsymbol{h}}(\boldsymbol{h}) = \beta\left(\sum_{i=1}^{m}\frac{(h_i-\beta v_i)^+}{\beta}(\boldsymbol{e}_i + \boldsymbol{v}) + \underbrace{\left(1 - \sum_{i=1}^{m}\frac{(h_i-\beta v_i)^+}{\beta}\right)}_{\geq 0}\boldsymbol{v}\right) \in \hat{\mathcal{U}}$$

where

$$\hat{\mathcal{U}} = \beta \cdot \mathsf{conv}\,(\boldsymbol{v}, \boldsymbol{e}_1 + \boldsymbol{v}, \ldots, \boldsymbol{e}_m + \boldsymbol{v})$$

Hence $\hat{\mathcal{U}}$ is a dominating set.

□

E.11 Domination for the generalized budget set

Propostion E.11.1. Let consider

$$\hat{\mathcal{U}} = \mathrm{conv}\left(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \frac{1}{m-1-2\theta}\boldsymbol{e}\right) \tag{E.8}$$

The set (E.8) dominates the uncertainty set (4.27).

Proof. Consider the uncertainty set (4.27) given by

$$\mathcal{U} = \left\{\boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} h_i \leq 1 + \theta(h_i + h_j) \;\; \forall i \neq j\right\}$$

and

$$\hat{\mathcal{U}} = \mathrm{conv}\left(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \frac{1}{m-1-2\theta}\boldsymbol{e}\right).$$

Note that in our setting we choose $\theta > \frac{m-1}{2}$. Take any $\boldsymbol{h} \in \mathcal{U}$. Suppose WLOG that

$$h_1 \leq h_2 \leq \ldots \leq h_m$$

Hence, by definition of $\mathcal{U}$

$$\boldsymbol{e}^T\boldsymbol{h} \leq 1 + \theta(h_1 + h_2)$$

To prove that $\hat{\mathcal{U}}$ dominates $\mathcal{U}$, it is sufficient to find $\alpha_1, \alpha_2, \ldots, \alpha_{m+1}$ non-negative reals with $\sum_{i=1}^{m+1} \alpha_i \leq 1$ such that for all $i \in [m]$,

$$h_i \leq \alpha_i + \frac{1}{m-1-2\theta}\alpha_{m+1}.$$

We choose $\alpha_{m+1} = (m-1-2\theta) \cdot \frac{h_1+h_2}{2}$, $\alpha_1 = h_1$ and for $i \geq 2$, $\alpha_i = h_i - \frac{h_1+h_2}{2}$. We can verify that

$$\alpha_1 + \frac{1}{m-1-2\theta}\alpha_{m+1} \geq \alpha_1 = h_1$$

and for $i \geq 2$,

$$\alpha_i + \frac{1}{m-1-2\theta}\alpha_{m+1} = h_i$$

Moreover, $\alpha_{m+1} \geq 0$, $\alpha_1 \geq 0$ and for $i \geq 2$, $\alpha_i \geq 0$ since $h_1 + h_2 = min_{i \neq j}(h_i + h_j)$. Finally,

$$\begin{aligned}\sum_{i=1}^{m+1} \alpha_i &= \sum_{i=1}^{m} h_i - (m-1) \cdot \frac{h_1 + h_2}{2} + (m-1-2\theta) \cdot \frac{h_1 + h_2}{2} \\ &\leq 1 + \theta(h_1 + h_2) - (m-1) \cdot \frac{h_1 + h_2}{2} + (m-1-2\theta) \cdot \frac{h_1 + h_2}{2} = 1.\end{aligned}$$

Note that the construction of this dominating set is slightly different from the general approach in Section 3 since we do not scale the unit vectors $\boldsymbol{e}_i$ in $\hat{\mathcal{U}}$.

□